Statistical learning and syntax: What can be learned, and what difference does meaning make?[1]

John N. Williams and Patrick Rebuschat

Many contributors to this volume have discussed statistical learning in the context of the word segmentation problem, stressing the role of contingencies between forms as an important cue to locating the boundaries between words. Having established the power of this kind of learning mechanism, an important question is whether it can be extended into other aspects of language learning. Here we focus on the learning of word order regularities, focusing primarily on adult second language (L2) learning. Is this just another form of sequence learning? Just as potential word boundaries can be induced through a statistical analysis of the contingencies between syllables, can syntactic structure be induced through an analysis of the contingencies between words?

An obvious place to look for research that is related to this issue is within the literature on artificial grammar learning (AGL) involving sequences of letters generated by finite-state grammars, or serial reaction time (SRT) tasks involving sequences of positions of a single stimulus. This is referred to generally as "implicit learning" research (see Cleeremans, Destrebecqz, & Boyer, 1998 for overviews; Perruchet, 2008; A. S. Reber, 1993; Shanks, 2005, for overviews). Following Misyak et al., and Dienes (this volume) we shall regard statistical learning and implicit learning research as essentially tapping a common underlying statistical learning mechanism. In both cases learning is assumed to occur non-intentionally, without explicit hypothesis formation or testing, and certainly without conscious access to the kinds of computations that appear to produce the learning effects (e.g. calculations of transition probabilities, or some similar statistic). Clearly this assumption is more obviously valid when not only the learning process, but also the resulting knowledge is shown to be implicit, as is often claimed in AGL experiments. If people do not know what they have learned then it seems highly unlikely that they would be aware of how they learned it.

1. We would like to thank Phillip Hamrick and Rebecca Sachs for helpful comments and Michelle To for generating our Matlab figures.

Within the AGL tradition an important issue has been whether people learn the underlying structure of an artificial grammar. If they do then their knowledge should generalise to new surface forms, just as knowledge of natural language grammars generalises to an infinite number of sentences. Reber (1969) investigated the effect of changing the lexicon or the grammar of the artificial language during the learning episode. The study consisted of two parts. The first part was a learning task that required subjects to memorize strings generated by a finite-state grammar. The second part consisted of a transfer task. Subjects were asked to continue memorizing letter sequences, but the letter sequences were modified without warning. For one group of subjects, the sequences were made up of the same letters as those in the first part, but a different finite-state grammar was used to generate them (old lexicon, new rules). For another group of subjects, the rules were the same, but the actual letters used to represent the grammar were changed (old rules, new lexicon). Reber found that changing the rules had a disruptive effect on subjects' memorization performance, but changing the lexicon had no detrimental effect. The memorization advantage, observed in Reber (1967, Experiment 1) was maintained as long as the rules remained the same. That is, subjects could "transfer" the knowledge acquired while memorizing one set of letter sequences to the memorization of a different set of letter sequences, even though both sets featured different lexicons. Reber concluded that implicit learning results in an abstract representation of the structure displayed in the stimulus environment.

The transfer effect has been frequently replicated (Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gomez, 1997; Gómez & Schvaneveldt, 1994; Matthews, et al., 1989; Tunney & Altmann, 2001; Whittlesea & Dorken, 1993). Several studies have shown transfer to be limited to certain conditions (Berry, 1991; Berry & Broadbent, 1984, 1988) but its existence seems largely noncontroversial. What remains contentious, however, is how to explain the underlying process. Reber's (1967, 1969) initial assumption was that subjects acquire abstract, rule-based knowledge during AGL experiments (see Reber, 1989, p. 114). According to the abstractionist account, the mental representation established during AGL consists of a symbolic structure that is independent of the original surface form of the training materials (Manza & Reber, 1997; A. S. Reber, 1989; A. S. Reber & Lewis, 1977). The transfer effect is explained by assuming that subjects learn rules which capture the structure of the training stimuli and then use this knowledge when judging whether test stimuli follow the same rules or not (e.g. Knowlton & Squire, 1996; Manza & Reber, 1997). An alternative view is that they merely learn the local patterns of alternation and repetition that the grammar generates, known as its "repetition structure" (Tunney & Altmann, 2001). Although repetition structures are an abstract form of knowledge (Marcus, et al., 1999), and may indeed be relevant to learning certain phonological rules, they are not the kinds of abstract rules that many linguists believe underlie natural language grammars in general.

What the AGL tradition does tell us, however, is that people are at least able to implicitly learn chunks of surface form; i.e. frequently occurring bigrams and trigrams in the input strings (Johnstone & Shanks, 1999; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). AGL learning can be modelled by chunking mechanisms (Cleeremans & Dienes, 2008). This is supported by research showing failures to implicitly learn grammars that cannot be induced through chunking surface forms (Matthews, et al., 1989; Shanks, Johnstone, & Staggs, 1997).

An important difference between typical AGL experiments and natural language learning is that AGLs exclude meaning. Therefore, they might underestimate the power of implicit learning to induce linguistic rules. There have been previous studies of miniature natural-like languages where the words have clear referents and the sentences are used to convey coherent messages (e.g., Friederici, Steinhauer, & Pfeifer, 2002; Morgan-Short, et al., 2010; Mueller, Girgsdies, & Friederici, 2008; Mueller, et al., 2005). These studies do appear to show learning of word order and agreement rules. But it is not clear whether the learning processes at work were purely statistical because the participants were in a situation where they had to work out the system in order to perform the task; that is, they were exposed to the system under intentional learning conditions. The main focus of those studies was on the knowledge that was acquired, and the associated brain regions and ERP responses evoked when it was used, rather than the learning process itself. But since these studies did not employ incidental learning paradigms it is not clear to what extent the results are due to simple statistical learning. Another problem with typical artificial language studies is that participants have to learn a very small system, consisting of just a few lexical items (just 14 words in the studies cited above). This makes it difficult to construct strong tests of generalisation, which should involve novel sentences with new lexis. In order to show that people have acquired knowledge that is more abstract than just surface chunks or transition probabilities between words we have to show generalisation to sentences with new lexis.

A method for examining learning of natural language word order regularities

To overcome the problems identified above Rebuschat & Williams (in press) used a semi-artificial language in which English lexis was combined with the syntax of a natural language that was unfamiliar to the participants. This resulted in sentences with apparently "scrambled" word order that were nevertheless perfectly comprehensible, e.g., "A few months ago ranted Chris about the government's plans", "George repeated today that the movers his furniture scratched", "After the instructor a sword brandished, focused Brian more on his defensive stance". In the exposure phase participants made plausibility judgments on sentences, hence engaging in a meaning-focused task that did not require them to analyse word order patterns (e.g., "Rose abandoned in the evening her cats on planet Venus" is implausible). After this they were informed that "The scrambling of the previous sentences was not arbitrary but followed a complex system" and were required to perform a grammaticality judgement test (GJT). Because native language lexis was employed, these sentences could use entirely new content words, and hence tested generalisation of syntactic knowledge in a way that was not confounded by the familiarity of specific sequences of words. The status of the knowledge that was acquired was assessed for each judgment through subjective measures of confidence (not confident, somewhat confident, very confident) and source attribution (guess, intuition, memory, rule) (following Dienes & Scott, 2005). We assumed that judgements of low and moderate confidence and/or guess or intuition source attribution reflected implicit knowledge (see Dienes, this volume).

It must be stressed that this paradigm was developed in the context of studying the initial stages of L2 acquisition. One might envisage a situation in which the second language learner has acquired some vocabulary in the language, begins to derive meaning from sentences in the input, but does not actually explicitly analyse the structure of those sentences. Following models of bilingual lexical development we assume that foreign language words initially associate with the meanings of first language translation equivalents (Kroll & Stewart, 1994; Kroll & Tokowicz, 2005). We also assume that they inherit the grammatical properties of their translation equivalents, as shown, for example, by cross-language syntactic priming (Salamoura & Williams, 2007; Schoonbaert, Hartsuiker, & Pickering, 2007), and cross-language priming of grammatical gender (Salamoura & Williams, 2008). Hence, in an experimental context, even if one were to burden participants with the task of learning novel lexical items, these

Table 1. Descriptions and Examples of the Three Verb Placement Rules

Rule	Description	Examples
V2	Finite verb placed in second phrasal position of main clauses that are not preceded by a subordinate clause.	*Today bought John the newspaper in the supermarket.*
V1	Finite verb placed in first position in main clauses that are preceded by a subordinate clause	*Because his parents the newspaper in the supermarket bought, spent John the evening in his study.*
VF	Finite verb placed in final position in all subordinate clauses	*Peter repeated today that the movers his furniture scratched.*

would merely inherit semantic and syntactic information from their first language equivalents. Therefore one might as well use first language words themselves, relieve the participants of the burden of learning and processing a large number of novel lexical items, and exploit the size of the native language vocabulary to construct strong tests of generalisation.

Returning to the Rebuschat & Williams (in press) study, the GJT test items either followed syntactic patterns that had been encountered in the exposure phase, or new patterns that violated the word order rules. The rules at issue concerned German verb placement, which is conditioned by clause type and clause sequence. In the exposure phase there were three basic patterns, exemplified in table 1. Note that these are not simple word position rules since the structural patterns concern the ordering of phrases (e.g., *Usually drew Jack his clients in a realistic fashion* and *A few months ago ranted Jessica about the government's plans* are both V2 constructions even though the verb is the fifth word in the second sentence). In the test phase participants performed a GJT on sentences with new lexis that either obeyed these verb position rules or violated them (see table 2). A control group performed this GJT without any prior exposure (they were instructed to judge how likely they thought it would be that each sentence would be grammatical in the world's languages). None of the participants had any knowledge of German or languages with relevantly similar word order rules.

The first question was what participants would learn incidentally about the word order rules of the language. The possibilities were that they would learn (i) associations between surface forms (e.g. that *John* was preceded by *bought*), (ii) structural patterns defined over abstract representations of

Table 2. Grammatical and Ungrammatical Patterns Used in the Testing Set

Pattern	*Grammatical*
V2	Yesterday scribbled David a long letter to his family.
V2VF	Recently have his parents an accountant consulted.
V2-VF	Paul argued recently that the chairman the wrong figures displayed.
VF-V1	Because his children fairy tales loved, invented John many stories.
	Ungrammatical
*VF	Recently Jim the Boston Marathon in four hours ran.
*V3VF	Yesterday the guitar was by David smashed.
*V2-V1	Recently maintained David that abstained his father from unhealthy food.
*VF-VF	Because his son an instrument wanted, David with the music teacher talked.

phrasal types, (e.g. for the V2 structure 'Time Phrase – Verb – Subject – Object'), or (iii) generalised syntactic rules (e.g., verb in second phrasal position in single clause sentences). If (i) were the case then participants should show no discrimination between grammatical and ungrammatical structures because all test sentences involve new lexis. In the case of (ii) and (iii) there should be discrimination between grammatical and ungrammatical sentences. Because the grammatical sentences repeat syntactic patterns from training it is not possible in this case to say whether they have learned patterns or rules (but see below for an experiment that attempts this).

It was found that the group who had received exposure performed significantly better than the controls on the grammatical structures (71% versus 36%), but did not differ on the ungrammatical structures (47% and 51%). Thus, there was clear evidence that, at a minimum, the experimental group incidentally learned the syntactic patterns underlying the exposure sentences, which it must be emphasised in this case concern the sequences of phrases and not individual words. The failure to reliably reject ungrammatical structures may suggest that there was no learning of grammatical rules, assuming that rules enable a clear identification of what is ruled out, as well as what is ruled in. However, because participants may be able to identify cases where a rule applies, but be unsure about the grammatical status of cases where it does not, then this is not strong evidence for a failure to learn rules.

Analysis of the data in terms of awareness measures revealed that overall performance was significantly above chance even for judgements where participants claimed that they were using intuition, as well as for judgements where they said they were using rules. Thus there was a mixture of implicit and explicit knowledge of the target structures, consistent with work in artificial grammar learning (Dienes, this volume).

In one sense these results are inconsistent with AGL research, on the basis of which we might have expected only learning of surface chunks, and hence no generalisation to sentences with new lexis. Of course, this inconsistency only arises if we adopt naïve assumptions about the kinds of representations over which people are able to learn the structure of sentences. If the learning mechanism is able to operate over more abstract categories than word forms, for example underlying phrasal categories, then what we have here is a simple case of learning sequences of phrase types. This secures generalisation to sentences with new lexis in a way that does not occur in AGL experiments. Similarly, Kaschak & Glenberg (2004) showed that participants can incidentally learn a novel construction in their native language which then generalises to sentences with new lexis. And Hudson Kam (2009) in an experiment employing an entirely artificial miniature language showed that adults can learn the word orders associated with specific verbs as abstract patterns which generalise to new lexis. The Rebuschat & Williams (in press) study extends these findings to a situation where the abstract patterns are defined over phrasal types and not individual words. However, given the chance performance on ungrammatical structures we cannot claim that participants learned the rules governing how clause type and sequence determines word order patterns within clauses.

Can abstract rules be learned?

The potential for learning abstract grammatical rules was explored more fully by Williams & Kuribara (2008) and Williams (2010) in the context of Japanese word order regularities. English words were combined with Japanese word order and case markers to form sentences such as “John-ga Mary-ni ring-o gave” (John gave Mary a ring), sentences which with prior instruction on the meanings of the case markers are readily interpretable (-ga marks nominative, -o accusative, and –ni dative). The procedure involved presenting many such “Japlish” sentences, reflecting a variety of constructions, in an exposure phase where the participant's task was to judge sentences for semantic plausibility (e.g. "John-ga Mary-ni planet-o gave", 'John gave Mary a planet', would be implausible).

The grammatical focus of these studies was on two interrelated properties of Japanese syntax: scrambling and head direction. The canonical Japanese word order is S(I)OV in simple sentences, and S[S(I)OV]V in complex sentences (e.g. Mary-ga John-ga book-o stole that said, 'Mary said that John stole a book'). But it is possible to scramble constituents to produce other word orders. From the perspective of generative theory the principle is that scrambling is defined as movement in the direction opposite to the head direction (Saito & Fukui, 1998). Since Japanese is right-headed, movement to the left is possible, resulting in structures in which the verb is always clause-final, but other arguments can appear in different positions (e.g., OSV, ISOV). In complex sentences arguments can even be moved out of the embedded clause (e.g. OS[SV]V, Book-o Mary-ga John-ga stole that said). We asked if participants were only exposed to a subset of possible scrambled structures (e.g., involving object and adjunct movement) whether they would subsequently generalise to structures involving movement of other constituents (e.g., indirect objects). This would indicate learning of a generalised notion of scrambling. On the other hand, they might only accept instances of scrambling of a type that they have experienced in the input. In this case they could only be said to have learned specific syntactic patterns. Of course, it is also important to test whether scrambling is appropriately constrained. Therefore there were also ungrammatical test structures that violated the principle of scrambling because the verb was not in final position.

Participants received either 194 or 388 sentences in the artificial language in the plausibility task. They then performed a GJT on sentences containing new lexis. There was also a control group who performed the GJT with no prior exposure. The results provided no evidence for learning a generalised notion of scrambling, nor indeed of the head-final property of the language. Rather, what people appeared to learn were the structures that they were trained on – as abstract syntactic patterns. But they did not learn that scrambling could be applied to new constituents (e.g., indirect objects). Nor did they learn the rule that all clauses had to end in verbs; that is, the head-final characteristic of the language. Despite the fact that every clause that they had received in training had ended in a verb they were still not better than chance at rejecting complex structures in which an embedded clause did not end in a verb, e.g., *S[SVO]V and *S[OVS]V. Rejection of simple sentences that did not end in a verb was rather better, e.g. *SIVO, *IOVS. Thus, they may have learned that sentences should end with a verb, but this was not the same as knowing that all clauses have to end with a verb, as would be required if the head direction of the language had been acquired.

But the participants did not simply learn a restricted set of syntactic patterns. If this had been the case they would have reliably rejected sentences that did not conform to patterns they had encountered before. This was clearly not the case, since absolute acceptance rates of new scrambles were around 60% in all conditions (this rate was not affected by level of exposure), and acceptance of ungrammatical structures was around 40% for the participants who had received exposure. An alternative hypothesis is that participants were basing their decisions on the similarity between a particular test sentence and the underlying statistical structure of the corpus of exposure sentences, and that acceptance rates reflected the familiarity of structures with respect to this statistical knowledge base. But the computation of similarity must have been based on abstract linguistic categories, rather than word forms. This hypothesis was supported by connectionist simple recurrent networks (SRNS) that were trained on the input coded as grammatical categories (e.g., S, O, V) and reproduced the relative rates of acceptance of the different test structures (Williams, 2010).

The contribution of linguistic knowledge

One way of construing the learning process occurring in the experiments described above is as a form of sequence learning over abstract grammatical categories. But is this the same kind of sequence learning process that is operative in, say, the more familiar statistical learning experiments that look at learning sequences of syllables (Saffran, Newport, & Aslin, 1996)? Does the nature of what is tracked and counted make any difference? As Johnson (this volume) points out, the view that learning is driven by a statistical learning mechanism leaves open the nature of the representations over which it operates. On the other hand, it is possible that because the representations are now grammatical, or semantic, different, non-statistical, learning mechanisms come into play.

A few previous studies have sought to assess how the addition of grammatical and semantic information affects the learning process. Robinson (2005) compared the performance of the same individuals on AGL and incidental natural language learning (participants were exposed to a micro-language based on Samoan). He focussed on two areas: first, whether the two learning tasks would pattern in similar ways with respect to measures of individual differences (IQ, measures of aptitude, etc.), and second, whether effects of chunk strength (Knowlton & Squire, 1996) would be found in both cases. He found very different patterns of results for the two tasks. There were different patterns of correlations with individual differences, and whilst effects of chunk strength and grammaticality were obtained in AGL, no such effects were obtained for Samoan. Thus different learning processes appeared to be at work in AGL and incidental natural language learning.

In fact the addition of meaning may decrease the learnability of sequences. When presented as meaningless syllable sequences, repetition-based patterns like AAB (e.g. *wo-wo-fe*) are easily learned by adults and infants (Marcus, et al., 1999), but when the categories correspond to grammatical categories, e.g. *scavenge-listen-camel* (VVN) these simple patterns are not (Endress & Hauser, 2009). It appears that linguistic knowledge can actually make us blind to patterns that would otherwise be readily learnable. Endress & Hauser argue that this is because syntactic processing, and learning, is modularised. Presumably this means that it does not draw on the same sequential learning mechanisms that operate in other domains. Mueller et al. (2008) examined learning of Japanese syntax in a miniature system and found that participants were actually more likely to produce native-like ERP signatures when meaning was removed. They suggest that the demands of semantic processing may interfere with learning the grammatical regularities. In this case it is meaning that makes people blind to syntax.

On the other hand the presence of semantic information might increase learnability. Non-adjacent dependencies, for example the A-B association in AXB sequences, have long been a focus of attention in the statistical learning literature, and in general have proved difficult to learn (Newport & Aslin, 2004; see also Hay & Lany, this volume; Perruchet & Poulin-Charronnat, this volume). Yet Amato & MacDonald (2010) found that in a language involving meaning, the equivalent of nonadjacent dependencies were easily learned.

The extent to which grammatical and semantic knowledge is helpful may depend upon the naturalness of the regularities that need to be learned. The only exception to the pattern-blindness found by Endress & Hauser (2009) was when the sequence made syntactic sense, e.g. *baby-water-juggle* (NNV) and *clever-fragile-water* (AAN). They argue that if the input can be interpreted syntactically (even if not matching syntactic patterns in the native language) its underlying structure can be learned. The ready learnability of the syntactically possible is reminiscent of a study by Cleary & Langley (2007) who found evidence for retention of the word order of meaningless, but grammatical, sentences, but not for ungrammatical sentences. For example, 'Beautiful transportation sheds temporary plants' primed sentences with the same sequence of grammatical categories. However, ungrammatical strings like 'sour a kick clean balloon hard' did not. The findings from Endress & Hauser (2009) and Cleary & Langley (2007) suggest that for grammatical knowledge to be helpful it has to resonate with the novel input, otherwise it might have a detrimental effect.

In order to gauge the impact of linguistic knowledge and/or meaning on learning word order regularities we need to employ methods that allow direct comparisons between linguistic and non-linguistic instantiations of the same underlying sequence. Here we shall discuss two studies in which we have tried to do this, albeit using very different methods.

Comparing meaningful sentences and nonsense analogues

In the Williams (2010) study the impact of grammatical knowledge and/or meaning was assessed by replacing the lexical items in the Japlish sentences with nonsense syllables. Each grammatical category was substituted with a set of similar nonsense syllables. For example, grammatical subjects were randomly replaced with one of *si/se/sa/so*, objects with one of *pi/pe/pa/po*, and verbs with one of *ki/ka/ko/ku*. For wh-words, *what* became *fu*, *who* became *fe*, and *when* became *fa*. The complementiser *that* was replaced with *me*. So for example, the original Japlish sentence *Pilot-ga that runway-o saw* became *Se-ga pi-o ku*; *Writer-ga who-ni what-o handed?* became *Si-ga fe-ni fu-o ki*?; *Detective-ga suspect-ga that car-o stole that announced* became *Si-ga so-ga pe-o ko me ku*. During the exposure phase participants performed a probe recognition task on each string (e.g., *Si-ga fe-ni fu-o ki*? might be followed by the probe *fe-o*, and the correct answer would be 'no'). The test strings were again nonsense analogues of the original test items, and care was taken that the syllable sequences were unique. Participants were asked to indicate whether they thought the strings were generated by the same system as the exposure strings. The results from the nonsense analogue and the (388) Japlish exposure condition were then compared. The mean endorsement of grammatical items was almost equivalent (70% and 72% for the analogue and Japlish respectively), and although endorsement of ungrammatical items was slightly lower in the analogue than in Japlish (24% and 30% respectively), this difference was not significant. To evaluate the correspondence between the experiments at a finer level of detail the correlation was calculated over the mean acceptance rate for each type of test structure (there were 21 different types, with 4 sentences per type). The correlation between the acceptance rates in the two experiments was $r = 0.832$, $p < 0.001$, meaning that 69% of the variability in Japlish was accounted for by endorsement rates in the analogue. The correspondence between the results of the two experiments suggests that similar learning processes were at work.

However, the correlation between the results of the studies was not perfect. Such discrepancies that did exist could be traced to the role of meaning and linguistic knowledge in the learning of, and making judgments upon, Japlish. For example, the long distance scrambles OS[SIV]V and IS[SOV]V were endorsed relatively poorly in Japlish compared to the nonsense analogue. It is reasonable to attribute this to the unusual mapping between linear surface form and meaning that results from extraction from the embedded clause (e.g., *That book-o John-ga student-ga borrowed that claimed*, John claimed that the student borrowed that book). Thus, the presence of meaning in Japlish actually suppressed acceptance of a pattern that, as a pattern, was more highly endorsed in the analogue. Whether the effort after semantic interpretation actually suppressed learning, or just the judgement process, is not clear however.

Also, in Japlish, multiple wh- sentences such as *Who-ga what-o bought*? were endorsed at the same high level as their declarative counterparts such as *Mary-ga book-o read*, even though they were less frequent in the input. In contrast, the acceptance rate for the analogues of the wh- patterns was lower than the declaratives (e.g., The *fe-ga fu-o ko* pattern was endorsed at a significantly lower rate than the *si-ga pe-o ku* pattern), as would be predicted by their frequency. Given that the same *–ga –o* pattern is present in all cases the difference is likely to arise because in Japlish, wh-questions and declaratives are encoded in terms of the same underlying grammatical pattern (e.g., SVO). Wh-words and nouns are already known to be in some sense functionally equivalent. But in the analogue there is no prior reason to suppose functional equivalence between syllables of the f_ class and the s_, and p_ classes.

The results of the comparison between Japlish and the analogue therefore suggest two ways in which prior grammatical knowledge and meaning can impact on learning. First, the complexity of semantic interpretation may detract either from pattern learning, or reduce the acceptability of complex structures in a GJT, a result that is similar to Mueller et al. (2008) who also targeted Japanese. Second, the potential for a grammatical interpretation changes the representations that are counted by the statistical learning mechanism. Structures that on the surface appear to be different (e.g. *Who-ga what-o bought* and *Mary-ga book-o read*) reinforce the same underlying pattern because they share a common abstract coding (SOV). In this case the coding is interpretable, and hence linguistic knowledge facilitates acquisition of a structure that is of relatively low frequency in the input. However, it must be stressed that overall the close correspondence between Japlish and the analogue suggests that the underlying learning mechanism was similar in the two cases. In the case of the nonsense analogue it is plausible that a simple statistical learning mechanism that tracks sequential contingencies between syllables was at work. And indeed, the same kind of SRN that was used to model Japlish produced a remarkably close fit to these data ($r^2 = 0.96$). Although the fit obtained for the Japlish model was significantly lower ($r^2 = 0.83$, significantly different at $p < 0.05$ using a Fisher r to Z transformation) the high fit in the two cases suggests that a similar learning mechanism was at work, except that it operated over different kinds of representation.

Linguistic and non-linguistic serial reaction time tasks

If we regard grammar learning as a form of sequence learning then it could be argued that GJTs are a poor method for measuring it because they do not measure sequence processing as such. For instance, they do not tell us which aspect of a target sentence the participant found unacceptable. Nor do they show that sequence knowledge affects real-time processing, as opposed to some off-line reflective process. And of course, psycholinguistic work on sentence processing has long since abandoned the use of GJTs in favour of on-line reading time and event related potential (ERP) measures. Assuming that these performance effects are a reflection of the linguistic knowledge that is actually used in language comprehension and production, then research on language learning needs to adopt these kinds of measures as well (for examples see Amato & MacDonald, 2010; Friederici, Steinhauer, & Pfeifer, 2002; Kaschak & Glenberg, 2004). What we need, therefore, are on-line methods that measure processing of sequences in time. Furthermore, in order to gauge the impact of linguistic knowledge and meaning we also need to make comparisons with processing the same sequence in a non-linguistic task.

The standard paradigm for examining sequence learning in psychology is the serial reaction time (SRT) task (e.g., Cleeremans & McClelland, 1991) in which participants track the movement of a stimulus by hitting corresponding response keys. Unbeknownst to the participants the sequence of stimulus positions follows a regular pattern and, by comparing reaction times to sequences that do or do not follow this pattern, learning effects can be evaluated. The SRT has become the preferred method for examining implicit learning in recent years, superseding the use of the AGL paradigm, most especially in neuropsychological research given that it is simple to administer to patient groups. The use of this paradigm also reflects a growing recognition of the importance of sequence learning, not only in motor skills, but also in complex cognitive skills such as language (Lieberman, 2007; Ullman, 2004).

Because in an SRT task the regularity to be learned involves a sequence of screen positions, the stimulus itself can either have linguistic content or not, whilst keeping the underlying sequence, as defined over screen positions, constant. Instead of seeing a constant stimulus move around screen positions the stimuli can be words or phrases that comprise a meaningful sentence. Screen positions can correspond to grammatical categories of phrases so that the spatial sequence is an analogue of the grammatical sequence. The task becomes a cross between a segment-by-segment reading procedure and an SRT task. The methodology allows sensitive on-line measurement of sequence learning effects and direct measurement of the impact of linguistic meaning.

We have used this modified SRT procedure to extend the Rebuschat & Williams (in press) work on learning German word order. Working on the assumption that people encode the exposure sentences as sequences of grammatical phrases a sentence like *Yesterday evening ate John a pizza at the restaurant* would be learned as an underlying phrasal sequence TP S O Adj (where TP stands for temporal adverbial phrase, and Adj stands for adjunct). The sequences can then be instantiated in a linguistic or non-linguistic SRT simply by assigning screen positions and corresponding keys to grammatical categories. Figure 1 shows one of the two layouts used in the experiment. The linguistic version of the task used words as stimuli, and the non-linguistic version used a constant meaningless stimulus (we used the string ‘XXX’). In the linguistic version the task is akin to a phrase-by-phrase reading paradigm except that the position of the phrase on the screen is variable, and the response key changes accordingly.

In the linguistic version a plausibility decision was required after every string in the exposure phase. In the non-linguistic version there was no additional task. The exposure phase was followed by a surprise test phase in which participants tapped through more sequences but this time, after each one, had to indicate whether it conformed to any of the patterns

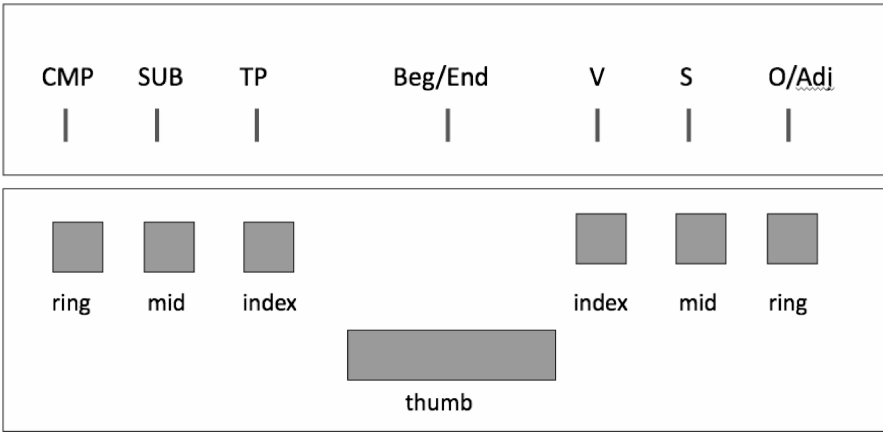

Figure 1. A possible layout of screen positions (top pane) and response keys (bottom pane) in the SRT experiments. Note that labels were not presented to the participants.
Note: CMP = complementizer, SUB = subordinator, TP = time phrase, V = Verb, S = Subject, O = Object, Adj = adjunct, Beg = Beginning of sentence, End = End of sentence.

seen in the exposure phase. In other words they were essentially performing a recognition memory task for sequence structures.

The test structures are shown in table 3. There were 5 examples of each type, or the same number of repetitions of the structures in the non-linguistic version. The exposure phase sentences comprised 40 examples of each of the grammatical structures (V2, V2-VF, VF-V1), or the equivalent number of sequences in the non-linguistic version. In the linguistic version different lexis was used in the exposure and test sentences.

The test sentences were formed into three sets in which the same words were configured into different word orders and rotated around conditions between participants. This enabled comparisons of reaction time profiles to be made in such a way as to control for lexical effects. A total of 24 participants were tested in the linguistic and non-linguistic versions respectively. None of them knew any German.

The reaction time effects in the SRT tasks were remarkably similar for the linguistic and non-linguistic versions. In both cases there were longer response times in the ungrammatical than the grammatical sequences at the points at which the sequences were violated. However, this was only the case for the short sequences/simple sentences; i.e. for the *V1 vs V2

Table 3. Example test items used in the SRT task

Pattern	*Grammatical*
V2	Yesterday scribbled David a letter to his brother.
V2-VF	Yesterday acknowledged David that in the evening Jenny the computer stole.
VF-V1	Because his dog many shoes devoured asked yesterday Peter the vet for advice.
	Ungrammatical
*V1	Scribbled yesterday David a letter to his brother.
*VF	Yesterday David a letter to his brother scribbled.
*V2-V2	Yesterday acknowledged David that in the evening stole Jenny the computer.
*VF-V2	Because his dog many shoes devoured, yesterday asked Peter the vet for advice.

and *VF vs V2 contrasts. By way of exemplification, figure 2 shows the reaction time profiles for the *VF vs V2 comparison for each version. No such slow-downs were evident in the long sequences. The addition of linguistic meaning did not appear to make the structure of long strings more learnable by this measure, but neither did it detract from learning of the short strings. Sequence learning was similar in the linguistic and non-linguistic versions.

Reaction times were not the only measure of learning in this experiment. Participants also made recognition memory decisions after each test string. The results are shown in figure 3. Discrimination between grammatical (equivalent in this context to grammaticality judgement, GJT) and ungrammatical structures is better for the non-linguistic version than the linguistic version. Mean endorsement rates for grammatical and ungrammatical structures are 61% and 42% in the non-linguistic version, but 68% and 63% in the linguistic version (the interaction is significant, $p < 0.05$). In fact, in the linguistic version only the V2 vs *V1 comparison is significant, $p < 0.001$. The contrast between the linguistic and non-linguistic versions is particularly striking for the *VF structure, which is readily discriminable from V2 in the non-linguistic version, $p < 0.001$, but actually has a numerically higher endorsement rate than V2 for the linguistic version. The addition of linguistic meaning obscured sensitivity to underlying sequential regularities by this measure.

In the case of the V2 vs *VF discrimination there is a clear dissociation between the GJT and RT data in the linguistic version. Within SRT research when recognition judgements show no effect of grammaticality but SRT performance does it is typically concluded that the sequential knowledge tapped by the SRT is unconscious (Norman, et al., 2007; P. J. Reber & Squire, 1998). This dissociation logic has been criticised on the grounds that the SRT may just be more sensitive than the judgement task (Shanks & Perruchet, 2002; Shanks, Wilkinson, & Channon, 2003). However, in the present case we also found the reverse dissociation – in the non-linguistic version there was discrimination between complex grammatical and ungrammatical structures in GJT but not in the SRT (see also P. J. Reber & Squire, 1999, for evidence of effects on recognition but not RT in an SRT task). We would argue, therefore, that the RT and GJT data are reflecting different kinds of knowledge.

If different knowledge sources are driving these tasks what is their nature, and what is their relevance to language learning? The SRT is commonly assumed to tap procedural learning. Given the similar results for the two versions, the simplest and most conservative interpretation is that in both cases the effects reflected learning in a "unidimensional" (Keele, et al., 2003) system that is only interested in the sequence as defined over screen position, whether the stimuli make linguistic sense or not. Learning was not affected by the presence of the secondary task of comprehension in the linguistic version, and at least in the case of *VF the RT effect seems to be implicit. Immunity to secondary task interference and implicitness are hallmarks of unidimensional learning according to Keele et al. (2003). Note also that learning here displays the limitations that one might expect of sequence learning in general; namely a sensitivity to the edges of short strings (hence rejection of *V1 and *VF) and an insensitivity to violations in the middle of long strings (Shukla et al, this volume).[2]

What kind of knowledge was tapped by the GJT? The obvious possibility is that in the linguistic version the GJT taps into linguistic representations of the sentences that are distinct from the underlying procedural knowledge reflected in the SRT. If we assume that, at this level, the material is represented in terms of sentences and clauses then the pattern of results makes sense. The acceptance of *VF can plausibly be regarded as an over-

2. Of course, this is not to say that long strings would not be learned with more exposure – connectionist simulations of this system (using an SRN) show that sensitivity to *V1 and *VF simply emerges before sensitivity to *V2-V2 and *VF-V2.

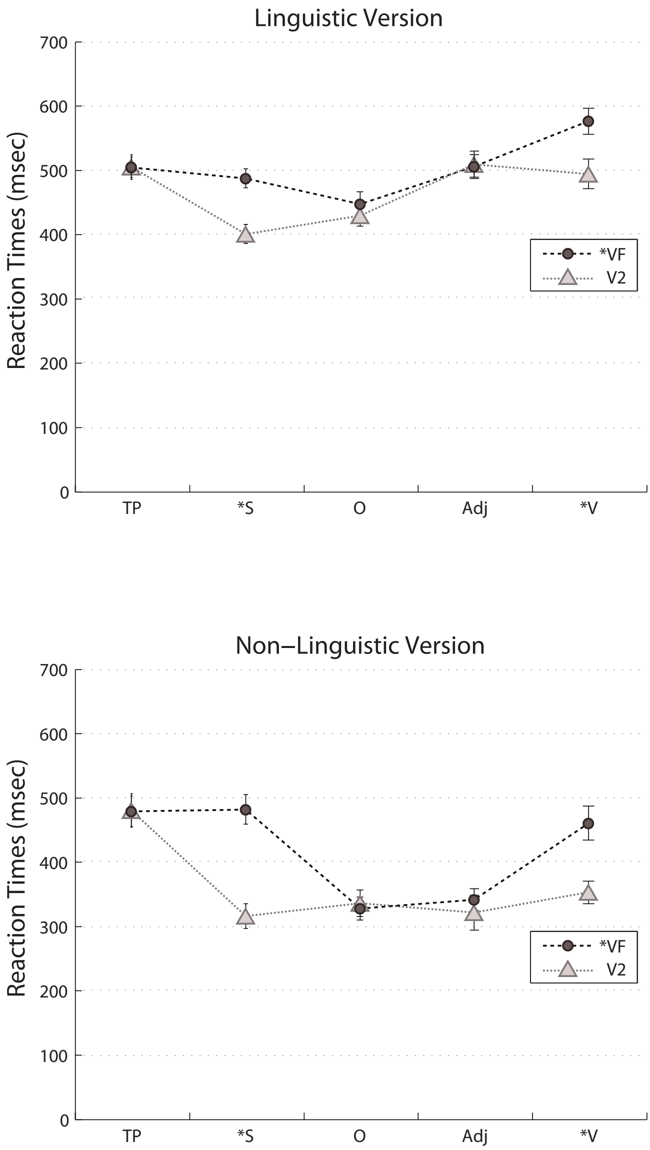

Figure 2. Reaction times in the SRT task for the V2 vs *VF comparison. The upper panel displays the results of the linguistic version and the lower panel the results of the non-linguistic version. In both graphs, the categories are arranged on the x-axis in the order in which they occurred in the ungrammatical sentences, and the categories that deviate from the expected grammatical V2 order are marked with asterisks. All differences between the V2 and *VF structures are significant at these positions. In each panel, errors are ± 1 S.E.M. based on data from 24 subjects.[3]

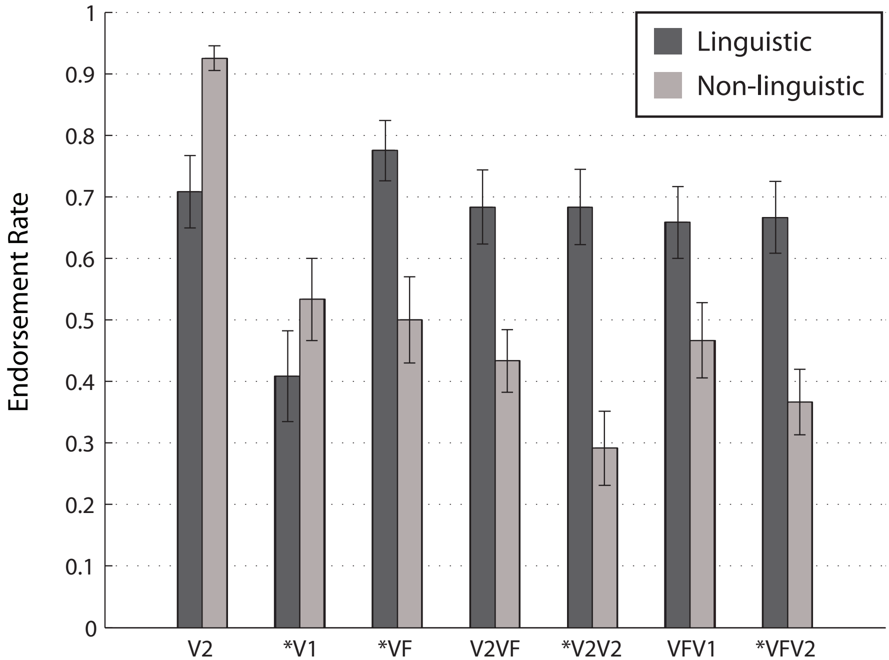

Figure 3. Endorsement rates in the trained groups. Errors are ±1 S.E.M. based on data from 24 subjects in each condition.

generalisation, based on the V2-VF structure, that *sentences* can end in verbs. And the elimination of discrimination between complex structures in the linguistic version plausibly reflects the fact that all complex sentences are composed of familiar clauses. Indeed, as figure 3 shows, the endorsement rates for the complex structures were uniformly higher in the linguistic version than the non-linguistic version ($p < 0.001$ across all complex structures). Clearly the participants had not learned the rules of clause type and

3. It could be argued that the differences between response times for a particular segment, e.g. V, arise because of positional effects. For example, the difference in RT for the verb could be because the verb is in second position in the V2 structure, but final position in the *VF structure. To evaluate this possibility linguistic and non-linguistic control conditions were run in which participants tapped through the test items with no prior exposure phase. There were no differences between grammatical and ungrammatical strings, showing that the differences shown in Figure 2 are effects of sequence knowledge, and not positional effects.

clause sequence, but if they had learned V2, V1 and VF as possible clause structures then all of the complex sentences would presumably have seemed fairly familiar. Therefore, at the level of knowledge tapped by the GJT, linguistic knowledge might obscure sensitivity to regularities that are apparent when the material is treated as pure sequence.

In terms of the learning mechanisms discussed in this volume, the GJT results in the linguistic version seem compatible with Perruchet & Poulin-Charronat's (this volume) unit formation view. They argue that associations between elements will not be learned unless they are held within the same attentional chunk. For example, in the domain of segmentation, prosody imposes an intonational phrase structure on the sound stream, and the edges of these units are a critical factor in learning sequential structure (Shukla, Nespor, & Mehler, 2007; Shukla et al., this volume). Associations will not be learned between elements that straddle an intonational phrase boundary (see Perruchet & Poulin-Charronat, this volume, for discussion). Likewise, it is not unreasonable to assume that in the larger language learning context, attentional chunks might be supported by meaning. What this means for our semi-artificial language is that the contingencies between elements at clause boundaries will be very hard to learn. And if beginnings and ends of strings are perceived as being edges of sentences (as high-level meaning-defined chunks) then the generalisation that sentences end in verbs will lead to acceptance of *VF.

This experiment demonstrates how procedural knowledge of sequences may be dissociable from other forms of knowledge, as tapped by different tasks. In the case of language, where syntactic and semantic analyses can lay on deeper and deeper layers of representation, this should come as no surprise. Nor should it be any surprise if the learning processes operating on these other layers of representation are not simply sequence learning processes of the type assumed to underlie, for example, learning to segment continuous speech.

Conclusion

In the studies reported here we have departed from the tradition of statistical learning and AGL research by exploring situations in which pre-existing grammatical knowledge and meaning can be brought to bear. We have defended this approach in the context of L2 acquisition on the assumption that L2 words tend to associate with the meanings and grammatical properties of their L1 translation equivalents, at least in the initial stages. The question we have posed is whether, when this rich linguistic knowledge is available, the nature or power of the learning process differs.

Clearly syntactic knowledge and meaning provide domains over which learning processes can operate that go beyond the kind of form-level learning traditionally examined in statistical learning research. As Johnson (this volume) remarks in the context of first language acquisition, "all computational models have to make some assumptions regarding what sort of information infants can perceive and process, and calculating statistics requires having some unit over which to do the calculations." Clearly, the adult L2 learner has access to a whole arsenal of linguistic categories that can be brought to bear on the input. And Sandoval, Gonzalez & Gomez (this volume), in the context of category learning suggest that "at that point learners could acquire phrase types characteristic of their native language where the content of the phrases would not be individual words (as in Saffran, 2001) but would be categories" – that is, having acquired lexical categories through statistical learning, then statistical learning can operate upon those categories to learn syntactic structure. Thus, in Japlish there was evidence that wh- and non-wh arguments, though different at the level of form were treated as equivalent, say at the abstract level of subjects and objects, in the meaningful version. And in the German SRT experiment, the distinction between 'long and short sequence', present in the non-linguistic version, was lost in the linguistic version where both were treated as 'sentences' (inviting an overgeneralisation of VF from complex sentences to simple ones).

If representational re-description alters the shape of the learning landscape, does it alter the nature of the learning mechanism? Is the learning process still essentially statistical – specifically one that is tuned to picking up contingencies between successive events (as might be modelled in a connectionist SRN for example)? In some cases, yes, it appears so. For example, Shukla et al. (this volume) provide evidence that lexical segmentation is learned by a process that is specifically tuned to consonants. But having imposed the relevant linguistic categorisation, the learning mechanism could still be regarded as essentially statistical. And in the present case, the fit of the Japlish real and analogue results to an SRN suggested that similar statistical sequence learning mechanisms were operating. Thus, the kind of process that tracks contingencies between syllables in the speech stream seems operative at the level of tracking contingencies between grammatical categories in sentences.

However, our German SRT study suggests that this assumption may be too simplistic, for here there was much less alignment between the linguistic and non-linguistic versions. In this case the GJT results for the linguistic version were more suggestive of chunking at the sentence and clause level, even though sensitivity to sequential structure was apparent on a procedural performance measure. What these results remind us of then is that multiple levels of representation may be formed simultaneously, tapped by different tasks, and that different learning mechanisms may be operative at different levels.

Our research also reveals limitations on what can be learned. The Japlish studies showed no learning of a generalised notion of scrambling, nor of head-final verb position as a generalised constraint. The German studies showed no evidence of learning the verb placement rules in terms of clause type and sequence. Rather, at best what the participants appeared to learn were just the syntactic patterns that they had received in training, and their responses to novel items in the grammaticality judgement tasks were determined by analogy to these patterns. These patterns were abstractly represented, hence supporting generalisation to sentences with new words (Endress, Nespor, & Mehler, 2009; Kaschak & Glenberg, 2004).

Of course, these limitations could be just a reflection of limited exposure. Although we have not examined the effects of prolonged exposure, we can make a prediction based on simulation work in Williams & Kuribara (2008). When the SRN simulation of Japlish was trained to asymptote (in terms of the error over training items) the only effect in the test was to enhance the response to trained items. The response to ungrammatical and novel grammatical test items did not change. This is presumably because, no matter how many times the network cycles through the training set, the statistical structure of the training data, on which the response to novel items is based, does not change. Thus, although the ability to discriminate old and new items improved, this was because of the greater familiarity of trained patterns, and not a reduced response to ungrammatical ones. Similarly, the effect of increased exposure to the German system would be predicted to have the effect of improving acceptance of trained patterns, but it would not improve rejection of ungrammatical ones. We predict therefore that the human ability to accept novel grammatical structures or reject ungrammatical ones will not change with increasing exposure because both are determined by analogy to examples encountered in the input. If performance on these structures does improve it must be because other learning processes, of a non-statistical kind, are engaged, for example explicit learning, or at the other extreme from statistical learning, a form of implicit learning guided by universal grammar.

Any discussion of the potential limitations of the statistical learning mechanism in relation to language must also be seen in the context of one's view of the nature of language. From a generative linguistic perspective, a failure to learn generalised notions of scrambling or head position would look like a fatal failing of the statistical approach. However, other views lay far more emphasis on the learning of syntactic patterns, rather than rules, such as the broadly emergentist (Ellis, 1998, 2002), usage-based (Lieven & Tomasello, 2008) and construction grammar approaches (see Ellis & O'Donnell, this volume; Goldberg, 2006). The nature of what is learned when we learn a first language continues to be hotly debated. But in L2 research there is a growing consensus that, at least in areas of grammar that are dissimilar to the L1, native-like competence or processing ability may not be achievable (Tolentino & Tokowicz, 2011). In this context we should not be perturbed by evidence for the limited power of statistical learning. This may in fact be an indication of the handicap under which the L2 learner is operating.

References

Altmann, G. T. M., Dienes, Z., & Goode, A.
1995 Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition*, 21, 899–912.

Amato, M. S., & MacDonald, M. C.
2010 Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, 116, 143–148.

Berry, D. C.
1991 The role of action in implicit learning. *Quarterly Journal of Experimental Psychology*, 43, 881–906.

Berry, D. C., & Broadbent, D. E.
1984 On the relationship between task performance and associated verbalisable knowledge. *Quarterly Journal of Experimental Psychology*, 36, 209–231.

Berry, D. C., & Broadbent, D. E.
1988 Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79, 251–272.

Brooks, R. L., & Vokey, R. J.
1991 Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 120, 316–323.

Cleary, A. M., & Langley, M. M.
2007 Retention of the structure underlying sentences. *Language and Cognitive Processes*, 22, 614–628.

Cleeremans, A., Destrebecqz, A., & Boyer, M.
1998 Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2, 406–417.

Cleeremans, A., & McClelland, J. L.
1991 Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235–253.

Dienes, Z., & Scott, R.
2005 Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, 69, 338–351.

Ellis, N. C.
1998 Emergentism, connectionism and language learning. *Language Learning*, 48, 631–664.

Ellis, N. C.
2002 Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.

Endress, A. D., & Hauser, M. D.
2009 Syntax-induced pattern deafness. *Proceedings of the National Acadamey of Sciences*, 106, 21001–21006.

Endress, A. D., Nespor, M., & Mehler, J.
2009 Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, 13, 348–353.

Friederici, A. D., Steinhauer, K., & Pfeifer, E.
2002 Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 529–534.

Goldberg, A. E.
2006 *Constructions at Work*. Oxford: Oxford University Press.

Gomez, R. L.
1997 Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, 33, 154–207.

Gómez, R. L., & Schvaneveldt, R. W.
1994 What is learned from artificial grammars? Transfer tests of simple associative knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 396–410.

Johnstone, T., & Shanks, D. R.
1999 Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 524–531.

Kaschak, M. P., & Glenberg, A. M.
2004 This construction needs learned. *Journal of Experimental Psychology: General*, 133, 450–467.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H.
2003 The cognitive and neural architecture of sequence representation. *Psychological Review*, 110, 316–339.

Knowlton, B. J., & Squire, L. R.
1996 Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 169–181.

Kroll, J. F., & Stewart, E.
1994 Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

Kroll, J. F., & Tokowicz, N.
2005 Models of bilingual representation and processing: looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism* (pp. 531–553). Oxford: Oxford University Press.

Lieberman, P.
2007 The evolution of human speech: its anatomical and neural bases. *Current Anthropology*, 48, 39–66.

Lieven, E., & Tomasello, M.
2008 Children's first language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 168–196). New York: Routledge.

Manza, L., & Reber, A. S.
1997 Representing artificial grammars: Transfer across stimulus forms and modalities. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford: Oxford University Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M.
1999 Rule learning in 7-month-old infants. *Science*, 283, 77–80.

Matthews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J.-R., & Druhan, B.
1989 The role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1083–1100.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T.
2010 Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, 60, 154–193.

Mueller, J. L., Girgsdies, S., & Friederici, A. D.
2008 The impact of semantic-free second-language training on ERPs during case processing. *Neuroscience Letters*, 443, 77–81.

Mueller, J. L., Hahne, A., Fujjj, Y., & Friederici, A. D.
2005 Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, 17, 1229–1244.

Norman, E., Price, M. C., Duff, S. C., & Mentzoni, R. A.
2007 Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition*, 16, 809–837.

Perruchet, P.
2008 Implicit learning. In H. L. Roediger (Ed.), *Cognitive psychology of memory. Learning and memory: A comprehensive reference.* (Vol. 2, pp. 597–621). Oxford: Elsevier.

Perruchet, P., & Pacteau, C.
1990 Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119, 264–275.

Reber, A. S.
1967 Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.

Reber, A. S.
1969 Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81, 115–119.

Reber, A. S.
1989 Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219–235.

Reber, A. S.
1993 *Implicit Learning and Tacit Knowledge*. Oxford: Oxford University Press.

Reber, A. S., & Lewis, S.
1977 Toward a theory of implicit learning: The analysis of the form and structure of a body of tacit knowledge. *Cognition*, 5, 333–361.

Reber, P. J., & Squire, L. R.
1998 Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, 10, 248–263.

Reber, P. J., & Squire, L. R.
1999 Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience*, 113, 235–242.

Rebuschat, P., & Williams, J. N.
in press Implicit learning of second language syntax. *Applied Psycholinguistics*.

Robinson, P.
2005 Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton &

Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, 27, 235–268.

Saffran, J. R.
2001 The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.

Saffran, J. R., Newport, E. L., & Aslin, R. N.
1996 Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.

Salamoura, A., & Williams, J. N.
2007 Processing verb argument structure across languages: Evidence for shared representations in the bilingual lexicon. *Applied Psycholinguistics*, 28, 627–660.

Salamoura, A., & Williams, J. N.
2008 The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, 10, 257–275.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J.
2007 The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, 56, 153–171.

Servan-Schreiber, D., & Anderson, J. R.
1990 Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 592–608.

Shanks, D. R.
2005 Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (pp. 202–220). London: Sage.

Shanks, D. R., Johnstone, T., & Staggs, L.
1997 Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50A, 216–252.

Shanks, D. R., & Perruchet, P.
2002 Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review*, 9, 362–367.

Shanks, D. R., Wilkinson, L., & Channon, S.
2003 Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology-Learning Memory and Cognition*, 29, 248–261.

Shukla, M., Nespor, M., & Mehler, J.
2007 An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54, 1–32.

Tolentino, L. C., & Tokowicz, N.
2011 Across languages, space, and time: A review of the role of cross-language similarity in L2 (morpho)syntactic processing as re-

vealed by fMRI and ERP methods. *Studies in Second Language Acquisition*, 33, 91–125.

Tunney, R. J., & Altmann, G. T. M.
2001 Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 27, 614–639.

Ullman, M. T.
2004 Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231–270.

Whittlesea, B. W. A., & Dorken, M. D.
1993 Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, 122, 227–248.

Williams, J. N.
2010 Initial incidental acquisition of word order regularities: Is it just sequence learning? *Language Learning*, 60, 221–244.

Williams, J. N., & Kuribara, C.
2008 Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, 118, 522–553.