The learnability of language

Insights from the implicit learning literature

Pierre Perruchet & Bénédicte Poulin-Charronnat
LEAD – CNRS UMR5022, Université de Bourgogne

The issue of the learnability of language contrasts the proposals of Chomsky (e.g. 1965), who claimed that the major part of language mastery involves innate domain-specific structures, to more recent nonnativist approaches, from the usage-based theories to Bayesian models, which contend that language acquisition rests on all-purpose domain-general learning processes. This chapter aims at examining the potential contribution to this issue of the literature on implicit learning, defined as the set of studies addressing the question of how participants learn in incidental conditions when they are faced with complex situations governed by arbitrary rules in laboratory settings. Overall, a striking parallelism emerges between usage-based approaches of language acquisition and implicit learning results, opening to a common research agenda.

The vast majority of theorizing about the learnability of language goes back to the earliest proposals of Chomsky (e.g. 1965), who claimed that the environment is not rich enough to allow a human learner to attain adult linguistic competence, an argument usually coined as the *poverty of stimulus* (*POS*) *argument*. As a consequence, language would be not learnable, and the major part of language mastery would involve innate domain-specific structures. This *nativist* claim has prompted lively debates, coming from various approaches that share the point that more is acquired from the input than initially thought (e.g. Bybee, 2007; Christiansen & Chater, 2008; MacWhinney, 2004; Seidenberg & MacDonald, 1999; Tomasello, 2003). For the sake of theoretical neutrality, these alternative approaches will be designed hereafter as *nonnativist* (see Ambridge & Lieven, 2011 for an overview of the contrasting approaches).

Using *nativist* and *nonnativist* to characterize the competing proposals is potentially misguiding. There is currently no one to posit that all aspects of language are innate (the diversities of languages in the world would be an obvious counterargument), and likewise there is no one to posit that all aspects of language are learned from experience (which would fail to explain that only humans master languages in

the first place). In fact, the core of the debates does not pertain to the precise location of the cursor on a continuous innate/learned dimension, but on a more qualitative distinction, namely whether language acquisition and especially syntax mastery rests on innate language-specific and modular mechanisms, or instead, on all-purpose domain-general learning processes. It has been rightly pointed out that this characterization eliminates from the landscape hybrid positions that would be logically viable,[1] but in keeping with most of the literature, and given that hybrid positions are not evoked in the following analysis, we use nativist (and nonnativist) to designate positions advocating for (or against) an innate language-specific module, which would be at the core of the language faculty.

This chapter aims at examining the potential contribution to these debates of the literature on implicit learning, defined as the set of studies addressing the question of how participants learn in incidental conditions when they are faced with complex situations governed by arbitrary rules in laboratory settings. Even though the rules underlying these situations do not mimic those of natural grammars, the ultimate objective of most researchers exploiting this methodology is to investigate how humans learn from natural settings, including first and second languages. The focus on implicitness still strengthens the a priori relevance of this research domain for language acquisition. Indeed, irrespective of the exact meaning given to the implicit/explicit distinction, there is no doubt that insofar as learning mechanisms are engaged in early language acquisition, they should be mostly implicit in nature. It is highly implausible that infants would have some intention to develop learning strategies, and would acquire explicit knowledge of the rules of their language.

In the following, we first outline some aspects of the debate on language learnability, then we examine how the knowledge gained from the studies on implicit learning in laboratory settings could influence this debate. Note that this chapter does *not* focus on the empirical studies that use natural languages as materials (mainly in the context of second-language acquisition) within a framework making heavy use of the concepts and methodology from implicit learning research. Excellent reviews on these studies are available elsewhere (e.g. Ellis, this volume; Williams & Kuribara, 2008; Williams & Rebuschat, 2012). Our objective is rather to emphasize the points of convergence existing between areas of research that primarily evolved in complete independence, with the postulate that conclusions emerging from independent sources of evidence may have add-on value.

---

1.   Newport (2011) notes that "the question of innateness and that of modularity are in principle distinct" (p. 281). See also Ambridge and Lieven (2011).

The debate on language learnability: A selective summary

The nativist position

We propose to decompose the logic of the nativists' arguments into a sequence of postulates or propositions. The following list is not intended to reflect a faithful historical sketch. For instance, proposition based on the lack of negative evidence (P5) does not come from Chomsky himself: Chomsky (e.g. 1965) initially assumed that child's input comprised "nonsentences", and the lack of negative evidence was first put forth *against* a nativist account of language acquisition. Moreover, we are aware that people defining themselves as nativist may disagree with one or another of the following points. However, we hope that the following sequence of propositions is fairly representative of the prevalent nativist claims.

(P1) The starting observation is that adults are able to produce and understand an infinite number of sentences in their language (and not only the small subset of utterances that they heard in the past) and to distinguish new grammatical sentences from ungrammatical sentences.

(P2) Accounting for this observation implies the mastery of a grammar. The characteristics of this grammar are determined by the properties of natural languages, and notably their hierarchical structure (by contrast with a purely linear organization). As an example, let us consider center-embedded linguistic productions. Suppose two sentences, a1b1 and a2b2, with *a* and *b* representing, respectively, NP and VP. The sentence a1a2b2b1 has a center-embedded structure because a2b2 is embedded within a1b1. An example of embedding in English is "the rat the cat ate stole the cheese", in which one relative clause ("the cat ate") is nested within the sentence ("the rat stole the cheese", see Figure 1). Note that the notion of hierarchical phrase structure is still fundamental for generativist linguists,[2] even though the initial notion of transformation (e.g. that declaratives and interrogatives are derived from a common kernel through transformational rules) is no longer advocated.

(P3) The objective of cognitive researchers is to account for the full knowledge of this grammar, coined as linguistic *competence*. Focusing on *performance*, which is the production of actual utterances affected by "such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors" (Chomsky, 1965, p. 3), is a misconception. Taking as illustration the center-embedded structure referred to above, producing a sentence with one or two levels of embedding attests to the capacity of producing an infinite number of embeddings recursively, and what has to be explained is the idealized mastery of the recursive rule. Recent

---

2. For the sake of simplicity, we focus here on a purely syntactic notion of linguistic competence, which is only a part of the generative school of thought.
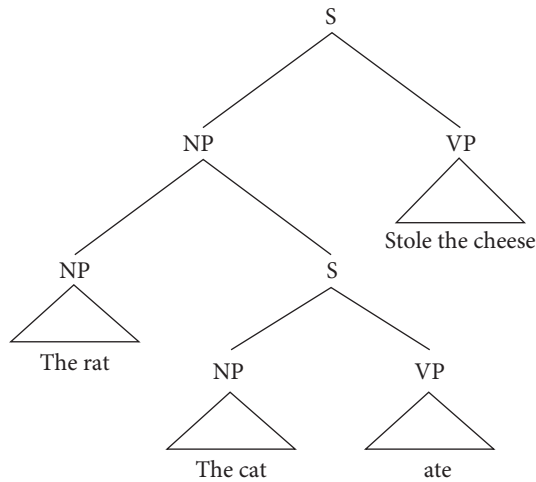
Figure 1. Schematic diagram of a phrase-structure grammar generating center-embedded sentences (triangles are used to mean that phrase internal structures are omitted).
S = Sentence, NP = Noun Phrase, VP = Verb Phrase

hypotheses make the central claim that the ability to process center-embedded structures is a critical cognitive feature distinguishing human from nonhuman cognition (Hauser, Chomsky, & Fitch, 2002).

(P4) Inferring such a grammar from experience is impossible. Indeed, any specific utterance is consistent with an infinite number of generalizations. The available evidence is not sufficient to eliminate irrelevant grammars, for two main reasons:

(P5) First, to infer that a grammatical rule is incorrect, infants should need *negative evidence* (i.e. information about what the language is not). For instance, to discover that center-embedding is the correct way for combining two sentences, infants should need to be told that utterances such as a1a2b1b2 are incorrect. To quote Pinker (1996), "without negative evidence, if the child guesses too large a language, the world can never tell him he is wrong" (p. xvii).

(P6) Second, the information provided through positive evidence (i.e. information about what the language is) is difficult to exploit through domain-general processes (such as associative learning) due to the problem of combinatorial explosion. To quote again Pinker (1996), "In most distributional learning procedures there are vast numbers of properties that a learner could record, and since the child is looking for correlations among these properties, he or she faces a combinatorial explosion of possibilities [...] To be sure, the inappropriate properties will correlate with no others and hence will eventually be ignored, leaving only the appropriate grammatical properties, but only after astronomical amounts of memory space, computation, or both." (p. 49).

(P7) Consideration of P5 and P6 implies that grammar knowledge can only be innate.

(P8) Innateness of the grammar implies in turn its universality, given that the grammar must be able to account for all languages around the world. Although the content of the Universal Grammar (UG) has substantially varied since the Chomsky's early writings, the postulate of UG remains as the cornerstone of the system (e.g. Chomsky, 2011).

Challenging views

Overall, challenging views focus on two complementary points. The nativist position would (1) overestimate the relevant linguistic knowledge and (2) underestimate the power of domain-general learning processes. To a large extent, these two points are in fact linked together, because the power of learning processes can only be assessed in relation to the to-be-explained observations. However, the relative weight of each factor differs between approaches. The Bayesian approach emphasizes the power of domain-general inferential processes, without questioning the postulate that language mastery is based on abstract grammatical knowledge. The starting point of disagreement of the other approaches, which are much more developed, is more related to the very nature of linguistic knowledge, with the renewed adequacy of domain-general cognitive process being in part a consequence of this change.

The Bayesian approach

The advocates of the Bayesian approach (e.g. Kemp, Perfors, & Tenenbaum, 2007) do not question the fact that infants must select the correct grammar over the space of all the possible grammars that could account for the actual input. In other words, P1, P2, and P3 remain unquestioned. However, instead of invoking innate constraints to guide hypothesis selection, they refer to an approach in which the choice is guided by the probability of the different grammars. In keeping with a Bayesian framework, this *posterior* probability would be dependent on both the *a priori* probability of the grammar, with the more compact grammar being the better valued, and the grammar's *likelihood*, defined as the probability for the current data to be generated by the grammar. There is a trade-off between these two components, because when the simplicity of the grammar increases, the fit to the data decreases, and conversely, precision is attained at the cost of additional complexity. The optimal balance is determined by the way of Bayesian inference.

Perfors, Tenenbaum, and Regier (2011) examined whether such a model exposed to a subset of the CHILDES database was able to correctly process the auxiliary-fronted interrogatives in English. They concluded that the hierarchical phrase-structure grammar favored by their model succeeded in this task, even though the few hours of

conversation entered into the model provided no direct evidence of auxiliary fronting in sentences comprising relative clauses.

Although this result is impressive,[3] it is worth stressing that the model requires considerable representational and computational power from the learners. Indeed, the model assumes that the learner is able to represent a hierarchical phrase-structure grammar among other possible grammars from the outset. Learning simply allows a learner without innate preference for this grammar to correctly select it over other grammars. Moreover, this selection requires heavy calculation. Perfors, Tenenbaum, Griffiths, and Xu (2011) acknowledge that "conventional algorithms for Bayesian inference running on conventional computing hardware take days or weeks to tackle problems that are vastly smaller than those the brain solves" (p. 314). We will come back to these points later.

Usage-based/emergentist approaches

The other alternatives to the nativist position no longer consider that language mastery implies knowledge of a full-blown grammar. In other words, the challenge concerns P2 and P3 of the nativist position. Linguistic performance replaces linguistic competence as a research objective. The prevalent perspective takes the form of so-called "usage-based approaches", which hold that children's language development emerges for the purpose of communicative intentions (comprehension and production).

The main tenet of usage-based approaches is that language development is item-based or exemplar-based. The early stages of language acquisition would be characterized by a collection of memories of previously heard utterances, each utterance being composed of one word, or more generally, of an unanalyzed chunk comprising a few words, associated with a meaning. These meaningful multiword linguistic units would be exploited both in comprehension and production.

For those familiar with the literature on the exemplar models of categorization (e.g. Medin & Schaffer, 1978) or implicit learning (e.g. Brooks, 1978), asserting the primacy of item-based units is misguiding, however. In fact, in usage-based approaches, the frequency of occurrence plays a crucial role in shaping the linguistic units (Bybee, 2007; for a review, see Diessel, 2007), so that usage-based models of language are more akin to statistical approaches than to the earlier line of thought focusing on the memory of exemplars. For instance, Bannard and Matthews (2008) selected frequently occurring chunks of four words in a corpus of child-directed language (e.g. *a drink of tea*), and they observed that 2-year and 3-year olds were more likely to repeat those

---

3.   Even if Berwick, Pietroski, Yankama, and Chomsky (2011) contest that such a demonstration rules out the POS argument. They notably contend that the grammar selected by Bayesian inference would fail to accommodate the full range of facts concerning constrained ambiguity.

sequences than to repeat control sequences matched in frequency with regard to individual words, but infrequent as a whole (e.g. *a drink of milk*). These and other results attest that children store frequent multi-word utterances as chunks.

There is evidence that the use of these rather specific units is not limited to the early stages of acquisition, and persists into adulthood. However, the prevalent idea is that these units also serve as a database for the construction of a lexicon, and more abstract grammatical regularities during subsequent stages of language development. The level of abstractness differs as a function of authors (to quote Ibbotson, 2013, p. 10,"there are almost as many usage-based theories as there are theorists"). At one extreme, several usage-based models hold that even simple lexical categories such as nouns and verbs, are not abstracted (e.g. Freudenthal, Pine, & Gobet, 2006), and that each item-based construction defines its own word class (Croft, 1991). Other authors accept the notion of lexical categories, and that fully articulated syntactic processing can take place in adult language on some occasions. However, the prevalent idea is that item-based generalizations are different from the very abstract structures envisioned in generative grammar, due to their local origin from a set of partially overlapping constructions. Instead of a fundamental division between lexical units and the rules that combine them, lexical items, word classes, and syntactic schemas are all viewed as constructions, which only differ along a continuum of specificity.

In this view, grammar emerges as generalizations from frequently used pattern, hence the term "emergentism" that is often used to designate this approach. As defined by MacWhinney (2010), the core difference between UG and emergentism is that "in the emergentist account, language depends on a set of domain-general mechanisms that ground language on the shape of the human body, brain, and society" (p. 472). As pointed by MacWhinney and others, statistical learning is one of the multiple domain-general mechanisms that serve to support language learning in this framework.

But for the argument of this chapter and given the available space, we refer the reader to available syntheses (e.g. MacWhinney, 2004) for more detailed information.

To sum up, the theories of language acquisition differ in striking ways. In the once prevalent nativist view (e.g. Chomsky, 1965), language acquisition proceeds through the selection of grammatical rules operating on abstract word classes, which are an instantiation of an innate and universal grammar. The task of researchers is to account for the mastery of this idealized grammar, called the linguistic competence. The Bayesian approach (e.g. Perfors et al. 2011) endorses the idea of abstract grammatical competence, but contends that humans are able to acquire this grammar from linguistic experiences through powerful inferential processes. Finally, in the now extensively developed usage-based approaches, the task of researchers is to account for the actual comprehension and productions of the learners, which take initially the form of item-based memories of specific utterances, followed by the extraction of some regularities through all-purpose learning processes.

The question is now: What are the implications of laboratory studies on implicit learning for this debate?

The implicit learning literature: A few key points and their implications

We examine below a few points that we consider as the key tenets of the implicit learning literature, and their implications with regard to language learnability. Although these points are largely consensual, some authors may take issue with specific aspects of our presentation. We take here for granted that implicit learning research and the more recent studies on statistical learning following the seminal Saffran and co-workers papers (e.g. Saffran, Aslin, & Newport, 1996) investigate closely related phenomena[4] (e.g. Misyak, Goldstein, & Christiansen, 2012; Perruchet & Pacton, 2006), and that these phenomena involve primarily the mechanisms that were identified in the conventional associative learning tradition (Perruchet & Poulin-Charronnat, 2012a). In addition, we pass over in silence the whole issue of whether the knowledge resulting from an implicit learning session is conscious or unconscious, even though this issue has raised a lively debate in the literature on implicit learning (e.g. for a review, see Shanks, 2005), and also in studies on second language acquisition (see Ellis, this volume). The main reason is that we are not sure that the very notion of "unconscious knowledge" makes sense (see Dulany, 1991; Perruchet & Vinter, 2002), and even if one endorses the validity of this notion, results in this area appear to be inconclusive and highly controversial. Space is lacking for providing a better-balanced handling of each point (for reviews on implicit learning, see Perruchet, 2008; Shanks, 2005). However, we hope that the remainder of this chapter will be suggestive even for the readers who do not share all our views.

No implicit rule abstraction

Certainly the most striking conclusion of implicit learning research is that when participants are faced with a complex situation governed by arbitrary rules in incidental learning conditions, they improve their performance, but without learning the underlying rules (e.g. Shanks, 2005). An immediate cautionary remark is in order: Formally, the subtitle of this section is an overstatement. The failure to observe implicit rule abstraction does not prove that implicit rule abstraction is impossible, but only that is has not yet been observed. However, two points are worth considering.

---

4.   We refer here to the stream of research on statistical learning that is now prevalent in the domain of language acquisition. It is clear that we do not deny the possibility of learning explicitly from distributional information (e.g. Onnis, 2012).

First, it can hardly be argued that this conclusion has been biased by researchers' theoretical entrenchment, which, for instance, would have led them to create ad-hoc arbitrary situations. Indeed, all the situations explored in this area of study have been initially devised to *demonstrate* rule abstraction, and they were initially claimed to be successful with regard to this objective. Without any exception, evidence that performance improvement in these situations was due to other mechanisms than rule discovery was provided in subsequent reappraisals.

Second, these reappraisals did not simply propose an optional, alternative account. The use of arbitrary situations allows manipulations that lead to definitely reject rule-based interpretations of the observed performance for these situations. By and large, these reappraisals are grounded on the fact that the rules selected by the experimenters lead to generate distributional biases in salient characteristics of the materials, and that participants learn from these correlated features. Experimental studies have regularly demonstrated either that keeping the rules while removing the correlated features impaired learning (e.g. Desmet, Poulin-Charronnat, Lalitte, & Perruchet, 2009; Wright & Burton, 1995), or that removing the rules while keeping the correlated features did not impair performance (e.g. Perruchet & Pacteau, 1990).

To illustrate, let us consider the artificial grammar learning paradigm, initially proposed by Reber (1967; see Pothos, 2007, for a review). Participants first study a set of letter strings generated from a finite-state grammar, such as shown in Figure 2. Participants are then subsequently informed about the rule-governed nature of the strings and asked to categorize new grammatical and nongrammatical letter strings. Participants are typically able to perform this task with better-than-chance accuracy. Reber's (1967) original proposal was that participants have internalized the finite-state grammar.
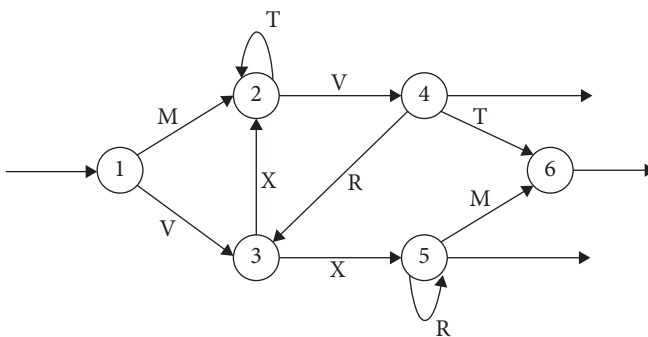
Figure 2. The artificial grammar used by Perruchet and Pacteau (1990) among others. For example, MTTV and VXVRXVT are grammatical, whereas MXVT is not grammatical

Looking at the grammar however, it appears that some associations between letters are possible (e.g. MV) and others impossible (e.g. MX), and that among the legal

associations, some are more frequent than others (e.g. RX presumably occurs more often than RM). If participants learn something about the frequency distribution of the pairs of letters (bigrams) that compose the study strings, they should perform subsequent grammaticality judgments better than chance. Perruchet and Pacteau (1990) tested this hypothesis. They reasoned that, if subjects learn only bigram information when faced with the whole strings, the direct presentation of the bigrams, which precludes the use of any high-level rules, should not change the final performance. The prediction was confirmed: The performance of participants who had learned using the complete grammatical strings (as usual) and those of participants trained using the bigrams from which these strings were composed were statistically indistinguishable.

Another argument against rule abstraction relies on the fact that in a given experiment, participants are exposed to a subset of the virtual full set of strings generated by the grammar. This subset cannot be perfectly representative of the full set for all aspects. For instance, the frequency distribution of the observed bigrams has a high probability of departing to some extent from the frequency of the bigrams composing the full set of strings. Reber and Lewis (1977) argued that if participants abstract the rules of the grammar, they should be sensitive to the bigram frequency of the virtual full set of strings, and not the frequency of occurrence of the bigrams composing the strings actually displayed in the study phase. Reber and Lewis provided empirical data supporting this hypothesis. The logic of the argument is indeed sound, but the supporting data turned out to be due to various methodological drawbacks. When these drawbacks were removed, participants were shown to be sensitive to the frequency distribution of the bigrams they actually perceived (Perruchet, Gallego, & Pacteau, 1992), running against the contention that studying grammatical letter strings gives access to the abstract structure of the grammar.

What does that mean for language learnability? At this point, certainly *not* that bigram information would be sufficient for mastering natural languages. However, if one takes for granted that even simple rules cannot be learned under incidental conditions, which are prevalent for language acquisition, it remains two possible conclusions: Either the rules of the language are given as a part of the human genetic endowment, such as claimed by nativists, or learners' linguistic knowledge is not rule-based as in the usage-based/ emergentist approaches. Noteworthy, this choice lets aside the Bayesian approach, in which rules are inferred thanks to powerful inductive processes. There is no sign of such processes in the implicit learning literature.

A shortcoming of the Bayesian approach could be to assume that learners proceed as linguists do when faced with a new language, by selecting the correct grammar among a host of possibilities. Note that this assumption is an explicit claim. For instance, Perfors et al. (2011) recurrently compare learners and linguists, as if the alleged parallelism strengthened the validity of the authors' assumptions. What is neglected here, however, is that there is a fundamental difference between a child and a linguist, namely that the former learns in implicit conditions, while the latter is driven

by intentional and conscious thoughts (in addition to having a much larger knowledge basis). In this regard, the Bayesian framework endorses an (often tacit) postulate that is construed by Perruchet and Vinter (2002) as the main flaw of the mainstream cognitive approach, namely taking the conscious mode of functioning of scientists as a model of one's unconscious processing (with the additional assertion that unconscious processing if free of the functional constraints of conscious thoughts).

The primacy of statistics: Implications for the lack of negative evidence

The preceding point should not be taken as a pessimistic assessment about the paucity of human learning abilities in incidental conditions. Instead, a second conclusion of implicit learning research is that participants quickly learn from complex settings, but they do so thanks to their powerful ability to exploit the statistical regularities available in the environment. Of course, statistical learning often provides only an approximation to rule-based knowledge, but the entire literature on implicit learning shows that this approximation is good enough to have misled researchers for years.

It is worth noting that acknowledging the primacy of statistical learning turns upside down the nativists' argument regarding the lack of negative evidence (P5). Negative evidence is arguably necessary whenever knowledge is conceived as rule-based, with learning consisting in hypothesis testing operations aimed at rule-selection. However, negative evidence is not only objectless, but harmful, whenever rule learning is no longer involved. This phenomenon has been clearly demonstrated in old studies on spelling. For instance, Brown (1988) explored the harmful effect of reading defective writings on subsequent spellings. He presented erroneous spellings of words to participants who were asked to rate how closely those misspellings resembled the correct spellings. The subsequent recognition of the correct spellings was impaired by the prior exposure to the incorrect alternatives. Thus, interference occurred even though participants were fully aware of the incorrectness of the writings they were exposed to (see also Perruchet, Rey, Hivert, & Pacton, 2006). The deleterious effect of being exposed to erroneous information on a subsequent memory test, even if this information is flagged as incorrect (as is the case for negative evidence), has been heavily explored, for instance in the context of pedagogical or testing methods involving multiple-choice procedures (e.g. Roediger & Marsh, 2005), and in studies comparing errorfull and errorless learning, especially in neuropsychological patients (e.g. Pitel, Perruchet, Vabret, Desgranges, Eustache, & Beaunieux, 2010).

From transitional probabilities to chunk formation

Language researchers from the nativist obedience generally postulate that statistical learning proceeds through the computation of pairwise associations. Pinker (e.g. 1996) asserts that a learner has to build a huge contingency table, the rows and columns of which are the properties of elements. For example, Pinker wrote that "he or

she could record of a given word that it occurs in the first (or second, or third, or *n*th) position in a sentence, that it is to the left (or right) of word X or word Y or …, or that it is to the left of the word sequence WXYZ, or that it occurs in the same sentence with word X (or words X, Y, Z, or some subset of them), and so on." (p. 49). In the more recent literature on statistical learning, the focus is on the computation of some kind of conditional or transitional probabilities between elements. At the computational level, this interpretation is generally implemented by connectionist networks, most often Simple Recurrent Networks (SRN).

The idea that the sensitivity to statistical regularities stems from the computations of conditional probabilities has been occasionally advocated in implicit learning research, but, by and large, this sensitivity is generally thought of as a by-product of chunk formation. The following example illustrates how the two accounts may differ in their predictions.

In a recent study by Endress and Mehler (2009), participants were exposed to an unsegmented artificial speech flow composed from trisyllabic words such as AB*X*, *Y*BC, and A*Z*C (with each letter standing for a syllable). These words were derived from the (unheard) "phantom-word" ABC. The phantom-word had exactly the same transitional probabilities between its constituent syllables (i.e. AB, BC, and A…C) than the (heard) trisyllabic words it generates. The reasoning was straightforward: If subjects have only learned pairwise relations, they should be unable to distinguish between the actual words and the phantom-words when both are played in a subsequent forced-choice test. However, if they have learned a word-like unit, that is some acoustical word candidates that could be later mapped as a whole to a meaning, they should select words over phantom-words. Although Endress and Mehler reported no difference between words and phantom-words, presumably due to the reduced level of intelligibility of their synthesized speech flow, Perruchet and Poulin-Charronnat (2012b) showed in three experiments using the same design that participants preferred words over phantom-words after only 5 min of exposure to the language. This result demonstrates that a chunk cannot be reduced to the concatenation of its parts, echoing the demonstration by Bannard and Matthews (2008; see above, 1.2.2.) that a frequent four-word utterance in child-directed language could not be reduced to the sequence of its components for 2-year and 3-year-old children.

Additional evidence for chunk formation stems from the discovery that adults (Perruchet & Desaulty, 2008) and infants (Pelucchi, Hay, & Saffran, 2009) are sensitive to *backward* transitional probabilities (given AB, the probability that B has been preceded by A) to the same extent as to forward transitional probabilities (the probability that A will be followed by B). This learners' ability is not compatible with the prediction-based logic of the SRNs. By contrast, the concept of chunk is naturally consistent with the bidirectionality of associations, and unsurprisingly, chunk-based models (e.g. Perruchet & Poulin-Charronnat, 2012b) account for the learners' sensitivity

to backward transitional probabilities without involving other mechanisms as those responsible for the sensitivity to forward transitional probabilities.

Conceiving chunks as the end-product of implicit learning (see Giroux & Rey, 2009, for further evidence) strengthens the bridge between implicit learning and usage-based approaches of language. Indeed, the starting point of usage-based approaches is the construction of multi-word units, conceived as unanalyzed chunks. One could object that chunks in implicit statistical learning investigations (1) are more similar to the words of the language than to the multi-word units of usage-based approaches, (2) do not include meaning or still, (3) link only within-modality elements. It must be realized, however, that learners cannot build units that are not present in the experimental materials, and participants have been typically exposed to random sequences of auditory or visual forms without any syntax or referents. There is evidence that if a phrase structure is present in an artificial language, the mechanisms that create words from syllables now create phrases from words, hence improving sensitivity to phrase structure (e.g. Thompson & Newport, 2007). Likewise, if one admits that chunking rests on associative learning mechanisms, there is no reason to restrict the chunks to a single modality (and for example to consider that cross-modal chunks need explicit learning). Indeed, as a rule, the stimuli used in most conditioning studies belong to different modalities (for a different view, see Ellis, this volume).

Beyond surface features

Implicit learning, as any form of associative learning, is often conceived as tightly linked to the surface features of the stimuli. The common opposition between rule-based and association-based frameworks leads to think that abstraction and association are mutually exclusive. This limitation would have devastating consequences. For instance, transposed to the issue of learnability, this would mean that associative processes could handle concrete events, such as the words "cat" or "eat", but not the abstract categories of noun and verb, and still less the linguistic constituents of higher levels.

It is worth stressing that the implicit learning literature shows unambiguously that associative processes operate on abstract entities. For instance, participants give evidence of learning about abstract features such as the distributional properties of the repetitions of nonadjacent letters in artificial grammar learning settings (Lotz & Kinder, 2006), the back-and-forth movement of a target in serial reaction time tasks (Perruchet, Gallego, & Savy, 1990), the frequency of octave intervals (Desmet et al. 2009), and the repetition of digits in invariant learning (Wright & Burton, 1995), among others. Acknowledging that implicit learning can handle abstract events is not a late and surreptitious addition to the conventional associative framework. The fact that complex representations can enter into associative links, and the high explanatory

power of this mode of functioning, has been pointed out in the modern literature on conditioning and learning (e.g. Mackintosh 1997, pp. 883–884).

Undoubtedly, acknowledging that associative learning processes can bear on abstract properties of the world considerably extends the power of associative mechanisms, but it should be stressed that this does not amount to assert that associative and rule-based mechanisms are similar. As cogently argued by Redington and Chater (2002), "surface-independence and rule-based knowledge are orthogonal concepts." (p. 124). Insofar as implicit learning processes are engaged, surface-independence is not an issue of concern, but rule-based operations on cognitive representations, whether concrete or abstract, have no place. This assessment is fully consistent with the evidence collected in a more linguistic context by Williams and collaborators (e.g. Williams & Kuribara, 2008; see also Williams & Rebuschat, 2012). Williams and Kuribara exposed native English speakers to a semi-artificial language combining the (unknown) syntax of Japanese with their familiar English lexicon in incidental conditions of learning. They showed that participants learned more than the specific word sequences they were exposed to, and were sensitive to the canonical word order of Japanese. Abstract coding was attested by the participants' ability to reliably accept sentences that contained different words from training sentences. However, they failed to acquire generalized grammatical rules, as shown, for instance, by their failure to reject ungrammatical sentences when their structure was complex.

It remains that jointly considering the fact that implicit learning results in the formation of chunks, and that the content of these chunks can include abstract components, fills a part of the gap separating a simplistic notion of statistical learning and the notion of hierarchical phrase-structure grammars. Indeed, the notion of transitional probability, or more generally the notion of pairwise association or sequential dependency, seems better fitted with linear grammars governing the position and the immediate succession of specific elements. By contrast, construing chunks as the end-product of implicit learning processes naturally opens to the possibility of a hierarchy of chunks, with superordinate and subordinate elements (Servan-Schreiber & Anderson, 1990). Moreover, accepting abstract representations as components makes it possible to conceive of these components, for instance, as lexical categories. A provisional conclusion is that, overall, the end-products of implicit learning seem much closer to a hierarchical structure such as represented in Figure 1 than could be anticipated.

The role of attention

Attention is a necessary and sufficient condition for implicit learning
In implicit learning settings, all the elements of the situation are overtly supraliminal. This does not entail that they are attended to. A question of major interest is whether

performance improvement depends on the amount of attention paid to these elements during the familiarization phase. The main strategy consists in adding a concurrent secondary task during the training session, then observing whether performance improvement is equivalent to that observed in a standard procedure. Although a few earlier implicit learning studies claimed that at least some forms of learning do not require attention, the bulk of recent evidence supports the opposite conclusion (for a review see Perruchet, 2008). As a rule, performances are degraded under double-task conditions. Without at least minimal attentional involvement, even simple and highly salient covariations or regularities turn out to be impossible to learn (e.g. Hoffmann & Sebald, 2005; Jiménez & Méndez, 1999). Conversely, nonsalient covariations or regularities can be learned provided that they are some reasons to pay joint attention to the relevant events (e.g. Pacton & Perruchet, 2008).

In line with other authors (e.g. Frensch & Miner, 1994; Logan & Etherton, 1994; Stadler, 1995), we consider here that associative learning is an automatic process that links together all of the components that are present in the attentional focus at a given point. It is worth noticing that this position, bypassing terminological issues, does not depart from the behaviorist view of Thorndike (e.g. 1932). Thorndike posited that a necessary condition for the creation of associative links is that the to-be-associated events are perceived as belonging to the same unit, a condition that he coined as the *principle of belonging*. Endorsing this framework is crucial with regard to the issue of learnability, because, as detailed in turn just below, the need for attention prevents from the combinatorial explosion predicted by the nativists, due to its selective character, while, at the same time, releasing the nature of the formed associations from unwanted constraints (such as event contiguity).

Implications for the combinatorial explosion

The combinatorial explosion refers to the exponential increment of possible associations between the available properties of the environment with the addition of new properties (P6). Because selectivity is a ubiquitous attribute of attentional processes, limiting the properties of the environment to those that are attentionally processed provides a straightforward response to the combinatorial explosion argument. However, this quantitative consideration is only a part of the matter. More importantly, it is worth stressing that attention does not act as a blind filter that would operate a random selection among the possible candidates for entering into new associations. Indeed, attention is naturally oriented towards events that have high chance of being relevant.

To illustrate, let us consider a question raised by Karmiloff-Smith (1992, p. 40) as an example of the general issue of learning form-meaning connections. When an adult points to a cat and says, "Look, a cat," how can the child pair the word "cat" with the whole animal, rather than, say, with the cat's whiskers, the color of the cat's fur, or

the background context? The question may find a response in the idea that infants' attention is captured by an array of stimuli sharing specific properties. Two of these properties are novelty and movement. If, at a given moment, several primitives are new for the infants, and/or move concurrently, it is highly probable that these primitives are processed conjointly in the attentional focus, hence forming a new unit. Now, if several primitives are new for a subject and/or move concurrently, there is also a good chance that they will be the components of one and the same meaningful unit, such as an actual object. Regarding the Karmiloff-Smith's problem, considering the auditory input first, "cat" is newer than "look," because "look" has been associated with many contexts before. As a consequence, it is highly probable that "cat" rather than "look" enters into the momentary attentional focus. On the other hand, it is also highly probable that the infant's attention is focused on the animal, which moves as a whole, rather than on one of its parts, or on the other elements of the context, which are presumably more familiar and motionless. In addition, infants' attention is also known to be guided by social cues such as adults' gaze direction (e.g. Wu & Kirkham, 2010), and it is highly likely that an adult saying "look a cat" is also looking at the cat rather than at the surrounding events.

To conclude, the apparent problem posed by the unmanageable number of potential units that can be initially perceived finds a simple solution thanks to the fact that attention is naturally captured by a tightly defined set of potentially relevant events. Note that attentional capture can be easily explained in evolutionary terms: Something new, moving, and capturing adults' attention has high chance of being relevant for survival.

Implications for learning nonadjacent dependencies

Undoubtedly, linguistic structures embed remote dependencies. That is to say, a relation exists between two events, A and C, irrespective of the intervening events (this structure is usually referred to as AXC, where X stands for a variable event, which is statistically independent of both A and C). Such relations are found, for instance, in morphosyntactic relationships (e.g. between auxiliaries and inflectional morphemes, as in "*is* writ*ing*," irrespective of the verb stem) and, crucially for the present debate, in hierarchical phrase structures (e.g. in the sentence: "the rat the cat ate stole the cheese", *the rat* (A) is associated with *stole the cheese* (B) irrespective of the relative clause *the cat ate* (X), which may be deleted or replaced by an infinite number of subordinate clauses, see Figure (1).

Accounting for nonadjacent dependencies within a statistical associative-based framework is a priori difficult. Undoubtedly, most experimental studies on learning have focused on the human abilities to detect and exploit the relations between adjacent elements. Moreover, from the "theory of contiguity" of Guthrie to the accounts of complex learning relying on the notion of chunks, standardly defined as the grouping

of a small number of contiguous events, the main theories of learning turn out to be devised for situations in which the relevant events are adjacent. A theory of associative learning positing that joint attention on the relevant events is both a necessary and a sufficient condition for learning associations provides a forceful argument for extending the relevance of associative learning principles to the learning of remote dependencies. This account is fully compatible with the conventional emphasis on the condition of contiguity, because the mental content composing the attentional focus at a given moment has a high chance of representing events that are close on spatial and/or temporal dimensions in the environment. However, the attentional content can also encompass events that are not adjacent in the environment, all simply because there are no functional or structural constraints linking human attention to spatially or temporally contiguous events. Each of us commonly merges present and past events in his/her attentionnal focus. An association may be composed of spatially or temporally remote events, provided that there is some reason for those events to be attentionally processed at the same time.

All-purpose mechanisms

The fact that implicit learning research is not targeted on language makes possible the use of multi-modal and multi-domain materials. Several studies have shown that, unsurprisingly, fine-grained results may differ according to sensory modality (e.g. Conway & Christiansen, 2005, for a comparison between touch, vision, and audition). More importantly for our concern, however, there is no evidence that linguistic stimuli (whatever their sensory modality) are learned differently from nonlinguistic stimuli. For instance, although most studies in artificial grammar learning involve consonant letters, a large variety of other stimuli have been used occasionally, such as geometric forms (e.g. Altmann, Dienes, & Goode, 1995), colors (Witt & Vinter, 2012), and sounds differing by their timbre (Bigand, Perruchet, & Boyer, 1998) or their pitch (Altmann et al. 1995), without noticeable difference. Still more interestingly, Williams and Rebuschat (2012) observed similar effects when directly comparing linguistic and nonlinguistic versions of serial reaction time tasks. Likewise, data on word segmentation have been successfully replicated with tones (Saffran, Johnson, Aslin, & Newport, 1999) or timbres (Tillmann & McAdams, 2004) instead of syllables.

The lack of substantial difference between linguistic and nonlinguistic stimuli in implicit learning research is certainly consistent with nonnativist approaches, which all rely on domain-general processes, but it is a double-edged sword: It could be taken as indicating that this literature taps only some low-level and uninteresting processes that would be common to language and other domains, but would be irrelevant to the deep issue of language learnability. The final question is: Is language so unique as nativists argue?

Considering the postulates described above as representative of a nativist position, the observation that adults are able to comprehend and to produce an infinite set of unheard sentences (P1) also applies to other domains, such as motor organization and music: Anyone can execute an action never performed in the past, anyone can feel a melody not previously heard as consistent with the rules of tonal music, and so on.

With regard to the problem raised by the lack of negative evidence (P5), Redington and Chater (1998, pp. 136–137) wrote:

> 'Almost all interesting learning from experience occurs without negative evidence, from finite sets of observations. For example, scientific theories are entirely grounded in observations of what *does* happen. Nonetheless, scientific progress seems possible. In learning about the physical structure of the world children, too, see only positive evidence. Yet they appear to learn a great deal about the world from this evidence alone. Since almost all interesting problems of learning from experience involve no negative evidence, and can manifestly be solved successfully, there seems no reason, at a general level, to assume that language learning from experience faces any special difficulties.'

Along this line, the combinatorial explosion that should result from the consideration of positive evidence alone according to nativists (P6) does not appear to be a language-specific issue. In fact, the very same question arises for conditioning in natural settings, in which animals have to select the appropriate conditional stimulus from the full array of stimuli available in the environment.

Although it is unquestionable that, at a fine-grained level, language is endowed with specific properties, it is not even sure that the main characteristic of the grammars on which nativist linguists focus is warranted. As claimed by Newport (2011), "hierarchical phrase structure, while supremely characteristic of human language, is clearly not unique to language […] and is utilized in descriptions of the complex organization of motor behavior, music, and many other domains" (p. 282). Overall, it appears that it is difficult to find identifiable properties that would be shared by no other human activities, or at the very least, the properties construed by nativists as specific to language – hence justifying particular considerations with regard to its acquisition – are much more general than once thought.

About learning failures: Is there an arbitrary/natural gap, after all?

Up to now, we have drawn a somewhat idealized picture of implicit learning research, as if any arbitrary situation coming from the experimenter's mind was quick and easy to learn. In fact, this picture is misguiding. Even in standard situations of artificial grammar learning with a finite state grammar, performances are far from perfect, with a ceiling around 60–70% correct responses. Moreover, complete learning failures occur on occasion. In particular, it appears that participants consistently fail to

learn in incidental conditions from materials generated by a so-called biconditional grammar (Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989; Shanks, Johnstone, & Staggs, 1997), at least when chunks of adjacent elements are controlled (Johnstone & Shanks, 2001). In a biconditional grammar, each element of the first part of a sequence is paired with the element in the corresponding location of the second part, generating strings such as *a1a2b1b2*. Why would *a1a2b1b2* be impossible to learn incidentally, while *a1a2b2b1*, the center-embedded structure shown in Figure 1, would be learnable?

This difference in learnability between an arbitrary structure and a natural structure, both of which appearing equally complex from an analytic standpoint, could give fuel to a nativist perspective, emphasizing the fact that we are genetically programmed to process the structures present in language. However, another general line of response could be that languages are the product of human culture, and that language properties are well-fitted to human learning abilities all simply because they reflect domain-general constraints that were present in *Homo sapiens* when languages emerged, and perhaps even in other primates (e.g. Christiansen & Chater, 2008; Tily, Frank, & Jaeger, 2011). Of course, what is true for language would be also true for other human productions such as music.

Rey, Perruchet, and Fagot (2012) provided a support to this idea from data collected in baboons. In their study, baboons were first exposed to six different pairs of visual shapes (noted a1b1, a2b2, …, a6b6) presented on a touch screen. They were trained to sequentially touch the first, and then the second element of each pair, while ignoring visual distractors. After this intensive paired associate training, they were shown two *a* elements in succession (e.g. a3 and a5), followed by the simultaneous presentation of *b* elements including b3 and b5. The question was: How did baboons order their responses? Results showed that baboons selected first b5, then b3, thus ordering their responses in keeping with a center-embedded structure rather than a biconditional grammar. It is worth stressing that the production of center-embedded structures was a *spontaneous* preference, and not the product of a supervised reinforcement, and that this preference, observed in baboons, cannot be accounted for by some transfer or generalization of a structure that would be primarily linguistic. The detailed reasons for this preference are not clear, but overall, it is likely that this mode of processing is induced by domain-general cognitive constraints, which would affect both action and perception (i.e. if baboons naturally produce b5b3 after seeing a3a5, they should be prone to associate a3 with b3 when exposed to a3a5b5b3). The final word is that, instead of considering that processing center-embedded structures is possible thanks to the innate endowment of UG, the considerations above suggest a radically different sketch: Both the presence of center-embedded structures in languages and their easy learning by humans would stem from a common cause, namely their dependence on ubiquitous domain-general cognitive constraints (see also Monaghan, 2011).

Needless to say, this sketch is nothing else than an oversimplified version of an account that should notably integrate, among other factors, the effects of cultural transmission on the evolution of languages, which could explain why what is only a probabilistic preference may end up as a universal pattern in language structure (Smith, 2012).

Conclusion

Summary and research agenda

Although rarely considered in the debate on language learnability, at least upon recent days, the literature on implicit learning in laboratory settings turns out to be a valuable source of information about how language could be learned. A major result is the ubiquitous failure to demonstrate rule abstraction, despite repeated attempts from researchers expecting a positive outcome. This result is inconsistent with a Bayesian approach. Indeed, insofar as laboratory situations are concerned, there is no evidence for implicit inferential processes that would allow participants to assess the Bayesian probabilities of multiple hypotheses about the rules governing the experimental settings. By contrast, the conclusions resulting from the implicit learning literature are fully consistent with other nonnativist approaches to language acquisition, based on the notions of emergentism and usage-based grammars.

It could be argued that, as a matter of principle, the conclusions derived from research on learning can only support an empiricist perspective. We do not think so. For instance, experimental participants could have proven to be unable to learn from complex arbitrary situations in incidental conditions, hence strengthening the need for a genetic endowment of natural language acquisition. The conclusion stays at the opposite, hence supporting the usage-based approaches, and we see as especially striking the convergence in evolution between two research domains that have evolved independently. If one lets aside the early attempt of Skinner (1957) to account for verbal behavior with basic learning processes, as well as the recent Bayesian models, the overall historical picture draws a shift from a rule-based perspective (irrespective of whether the rules are instantiations of an innate UG or inferred from the data) to a view focusing on the sensitivity to the statistical structure of the environment. This parallelism may stem from the fact that a rule-based account is much simpler to conceive and comprehend for the researcher than an approach in which the behavioral sensitivity to complex environmental structures is conceived as the product of domain-general associative mechanisms. As a consequence, it is logical that accounts based on domain-general learning mechanisms exploiting the statistical structure of the environment emerge only during a later stage of research development.

We surmise that this parallelism in evolution could extend into the future. Indeed, we feel that implicit learning research and usage-based/statistical approaches to language are faced to similar questions and hence, could share a part of their research agenda. In our view, the main point is the following. As mentioned above, the two fields of research share the view that the early phases of adaptation to a complex structured environment essentially consists in the formation of unanalyzed chunks of a few elements, which serve for both perception and production. This is a strong and worthwhile point of convergence. However, what is less clear is the follow-up of these chunks with extended exposure to the stimuli. Regarding natural language, certainly a part of them may remain functional throughout adult life, but it is hardly questionable that initial chunks also serve as a base for further abstraction. As claimed by Ibbotson (2013) in relation to usage-based theories, "clear developmental predictions about how the process of abstraction should develop, including which systems should become abstract first, are needed" (p. 9). Research on implicit learning in laboratory settings could help in this research project. How does abstraction proceed? How do learners draw analogies between chunks? What is the final state and in particular, is there a need for the formation of abstract categories (similar to noun and verb in language)? Is the intervention of conscious and explicit processes necessary at this level of analysis? These are some of the questions that should be addressed in the next years. This research agenda needs the elaboration of new experimental settings. Indeed, standard procedures of artificial grammar learning or serial reaction times are obviously unable to handle such issues, all simply because the materials do not include the relevant structures. Whether or not research on natural language and studies using arbitrary nonlinguistic events in laboratory will go on to bring out similar responses is open to speculation.

Is nativism definitely outdated?

In the introduction of this chapter, the nativist (and nonnativist) positions were defined in relation to their endorsement (or their rejection) of an innate language-specific module, which would be at the core of the language faculty. The literature covered in this chapter strongly suggests that there is no need for an innate language-specific module, even if there is still a long way to go before reaching a full-blown account of language mastery based on domain-general learning processes. The lack of project directly aimed at deepening the nativist/nonnativist debate in the research agenda above could be taken as indicative that we construe this issue as definitely settled in favor of nonnativist approaches. The matter is not so simple, however. It should be clear that such a definitive conclusion cannot be endorsed by nativists. Indeed, nativists adhere to the foundational postulate that cognitive researchers must address linguistic *competence* (P3). Undoubtedly, and in keeping with the ubiquitous failure

to demonstrate rule abstraction in laboratory situations, usage-based approaches to language acquisition do not account for linguistic competence.

Let us take as illustration of this issue the center-embedded structure referred to throughout this chapter. In principle, given that in spoken language, multiple center-embeddings even of degree two are practically nonexistent, the actual mastery of this structure raises no insurmountable problem for an empiricist approach. As claimed by Evans and Levinson (2009), "once recursion is quarantined to one level of nesting it is always possible to use a more limited type of grammar, such as a finite-state grammar, to generate it" (p. 443). Learning this structure appears all the easier that center-embedding is the natural way of integrating multiple associations even in non-linguistic contexts (Rey et al. 2012). However, for nativists, all of this has only minor interest if any, because, in their view, what has to be explained is the idealized mastery of the recursive rule that allows handling a potentially unbounded set of linguistic expressions.

Despite a number of conceptual changes throughout his ground-breaking career, Chomsky has persisted in claiming the importance of studying competence and its support: the UG. In a recent paper, Chomsky (2011) defines the language in reference to UG, asserting that opponents to this postulate argue in fact for the "nonexistence of language". Is this claim warranted? Our own response would be that calling for unlimited competence as a way of accounting for the production of language was certainly a reasonable option at a time where language looked as definitely out of reach of an empiricist account. Whether this option is still reasonable today is more questionable. It is worth reminding that a nativist position is not supported by positive arguments, such as some kind of direct evidence that UG is implemented into the human genetic endowment. The call for innate processes stems essentially from the alleged failure to account for language acquisition through learning processes (P7). Providing growing evidence that domain-general learning processes are in fact appropriate for language acquisition such as observed in children, and not affected by the limitations that have been traditionally construed as lethal (such as the lack of negative evidence, and the combinatorial explosion that would result from blind associative learning processes),[5] deprives a nativist account from most of its original motivations. In this context, still claiming that what has to be explained is an idealized competence, and that any performance improvement is driven by the full knowledge of a grammar, can certainly

---

5. It is not clear that all nativists are fully aware of these conceptual changes, given the rather limited knowledge of the associative learning literature they exhibit. For instance, for Gregg (2003), associative learning is the process whereby "given enough yellow bananas, one learns that bananas are yellow" (p. 120). No doubt that with such a view, associative learning appears underpowered to account for even the most basic linguistic productions, let alone linguistic competence.

not be proven to be wrong, but appears increasingly as reflecting nothing more than a dogmatic entrenchment.

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*(4), 899–912. DOI: 10.1037/0278-7393.21.4.899

Ambridge, B. & Lieven, E.V.M. (2011). *Child language acquisition: Contrasting theoretical approaches.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511975073

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*(3), 241–248. DOI: 10.1111/j.1467-9280.2008.02075.x

Berwick, R.C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science, 35*(7), 1207–1242. DOI: 10.1111/j.1551-6709.2011.01189.x

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 17*(3), 577–600.

Brooks, L.R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A.S. (1988). Encountering misspellings and spelling performance: Why wrong isn't right. *Journal of Educational Psychology, 80*, 488–494. DOI: 10.1037/0022-0663.80.4.488

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780195301571.001.0001

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language Learning and Development, 7*(4), 263–278. DOI: 10.1080/15475441.2011.584041

Christiansen, M.H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*(5), 489–558.

Conway, C.M., & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24–39. DOI: 10.1037/0278-7393.31.1.24

Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information.* Chicago IL: University of Chicago Press.

Desmet, C., Poulin-Charronnat, B., Lalitte, P., & Perruchet, P. (2009). Implicit learning of nonlocal musical rules: A comment on Kuhn and Dienes (2005). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 299–305. DOI: 10.1037/a0014278

Diessel, H. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology, 25*, 108–127. DOI: 10.1016/j.newideapsych.2007.02.002

Dulany, D.E. (1991). Conscious representation and thought systems. In R. S. Wyer, & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 4; pp. 97–120), Hillsdale, NJ: Lawrence Erlbaum Associates.

Endress, A.D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*, 351–367. DOI: 10.1016/j.jml.2008.10.003

Evans, N., & Levinson, S.C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(5), 429–494. DOI: 10.1017/S0140525X0999094X

Frensch, P.A., & Miner, C.S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition, 22*(1), 95–110. DOI: 10.3758/BF03202765

Freudenthal, D., Pine, J.M., & Gobet, F. (2006). Modelling the development of children's use of optional infinitives in Dutch and English using MOSAIC. *Cognitive Science, 30*, 277–310. DOI: 10.1207/s15516709cog0000_47

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33*(2), 260–272. DOI: 10.1111/j.1551-6709.2009.01012.x

Gregg, K.R. (2003). The state of emergentism in second language acquisition. *Second Language Research, 19*, 95–128. DOI: 10.1191/0267658303sr213oa

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579. DOI: 10.1126/science.298.5598.1569

Hoffman, J., & Sebald, A. (2005). When obvious covariations are not even learned implicitly. *European Journal of Cognitive Psychology, 17*(4), 449–480. DOI: 10.1080/09541440440000203

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psycholology, 4*(Article 255), 1–15.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 236–259. DOI: 10.1037/0278-7393.25.1.236

Johnstone, T., & Shanks, D.R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42*(1), 61–112. DOI: 10.1006/cogp.2000.0743

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: The MIT Press/Bradford Books.

Kemp, C., Perfors, A., & Tenenbaum, J. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*(3), 307–321. DOI: 10.1111/j.1467-7687.2007.00585.x

Logan, D.G., & Etherton, J.L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1022–1050. DOI: 10.1037/0278-7393.20.5.1022

Lotz, A., & Kinder, A. (2006). Transfer in artificial grammar learning: The role of repetition information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 707–715. DOI: 10.1037/0278-7393.32.4.707

Mackintosh, N.J. (1997). Has the wheel turned full circle? Fifty years of learning theory, 1946–1996. *The Quarterly Journal of Experimental Psychology, 50A*(4), 879–898.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child* Language, 31, 883–914. DOI: 10.1017/S0305000904006336

MacWhinney, B. (2010). Language development. In M. Borstein & M. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 467–508). New York, NY: Psychology Press.

Mathews, R.C., Buss, R.R., Stanley, W.B., Blanchard-Fields, F., Cho, J.R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1083–1100. DOI: 10.1037/0278-7393.15.6.1083

Medin, D.L., Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238. DOI: 10.1037/0033-295X.85.3.207

Misyak, J.B., Goldstein, M.H., & Christiansen, M.H. (2012). Statistical-sequential learning in development. In Rebuschat & Williams (Eds.) (pp. 13–54).

Monaghan, P. (2011). Design features of language emerge from general purpose learning mechanisms. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Boston, MA: Cognitive Science Society.

Newport, E.L. (2011). The modularity issue in language acquisition: A rapprochement? Comments on Gallistel and Chomsky. *Language Learning and Development, 7*(4), 279–286. DOI: 10.1080/15475441.2011.605309

Onnis, L. (2012). The potential contribution of statistical learning to second language acquisition. In Rebuschat & Williams (Eds.) (pp. 203–235).

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and non-adjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 80–96. DOI: 10.1037/0278-7393.34.1.80

Pelucchi, B., Hay, J.F., & Saffran, J.R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*(2), 244–247.
DOI: 10.1016/j.cognition.2009.07.011

Perfors, A., Tenenbaum, J.B., Griffiths, T.L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*(3), 302–321.
DOI: 10.1016/j.cognition.2010.11.015

Perfors, A., Tenenbaum, J.B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*, 306–338. DOI: 10.1016/j.cognition.2010.11.001

Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.), *Cognitive psychology of memory* (*Vol.2*): *of Learning and memory: A comprehensive reference* (pp. 597–621). Oxford: Elsevier.
DOI: 10.1016/B978-012370509-9.00149-2

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition, 36*(7), 1299–1305. DOI: 10.3758/MC.36.7.1299

Perruchet, P., Gallego, J., & Pacteau, C. (1992). A reinterpretation of some earlier evidence for abstractiveness of implicitly acquired knowledge. *The Quarterly Journal of Experimental Psychology*, *44A*(2), 193–210. DOI: 10.1080/02724989243000000

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology*, *22*(4), 493–516. DOI: 10.1016/0010-0285(90)90011-R

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*(3), 264–275. DOI: 10.1037/0096-3445.119.3.264

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233–238.
DOI: 10.1016/j.tics.2006.03.006

Perruchet, P., & Poulin-Charronnat, B. (2012a). Word segmentation: Trading the (new but poor) concept of statistical computation for the (old, but richer) associative approach. In Rebuschat & Williams (Eds.) (pp. 119–143).

Perruchet, P., & Poulin-Charronnat, B. (2012b). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language, 66*, 807–818. DOI: 10.1016/j.jml.2012.02.010

Perruchet, P., Rey, A., Hivert, E., & Pacton, S. (2006). Do distractors interfere with memory for study pairs in associative recognition? *Memory & Cognition*, *34*(5), 1046–1054. DOI: 10.3758/BF03193251

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences, 25*, 297–388.

Pinker, S. (1996). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pitel, A.L., Perruchet, P., Vabret, F., Desgranges, B., Eustache, F., & Beaunieux, H. (2010). The advantage of errorless learning for the acquisition of new concepts' labels in alcoholics. *Psychological Medicine*, *40*(3), 497–502. DOI: 10.1017/S0033291709990626

Pothos, E.M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*(2), 227–244. DOI: 10.1037/0033-2909.133.2.227

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863. DOI: 10.1016/S0022-5371(67)80149-X

Reber, A.S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition, 5*, 333–361. DOI: 10.1016/0010-0277(77)90020-8

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 12*(2–3), 129–191. DOI: 10.1080/016909698386500

Redington, M. & Chater, N. (2002). Knowledge representation and transfer in artificial grammar learning. In R. French & A. Cleeremans (Eds.), *Implicit Learning and Consciousness*. Hove: Psychology Press.

Rebuschat, P., & Williams, J. (Eds.). 2012. *Statistical learning and language acquisition.* Berlin: De Gruyter Mouton.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (Papio Papio). *Cognition, 123*(1), 180–184. DOI: 10.1016/j.cognition.2011.12.005

Roediger, H.L., & Marsh, E.J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1155–1159. DOI: 10.1037/0278-7393.31.5.1155

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. DOI: 10.1126/science.274.5294.1926

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. DOI: 10.1016/S0010-0277(98)00075-4

Seidenberg, M.S., & MacDonald, M.C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569–588. DOI: 10.1207/s15516709cog2304_8

Servan-Schreiber, E., & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 592–608. DOI: 10.1037/0278-7393.16.4.592

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, 50A*(1), 216–252. DOI: 10.1080/713755680

Skinner, B.F. (1957). *Verbal behavior*. New York, NY: Appleton. DOI: 10.1037/11256-000

Smith, K. (2012). Evolutionary perspectives on statistical learning. In Rebuschat & Williams (Eds.) (pp. 409–431).

Stadler, M.A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 674–685. DOI: 10.1037/0278-7393.21.3.674

Thompson, S.P., & Newport, E.L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1–42. DOI: 10.1080/15475440709336999

Thorndike, E.L. (1932). *The fundamentals of learning*. New York, NY: Teachers College, Columbia University. DOI: 10.1037/10976-000

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1131–1142. DOI: 10.1037/0278-7393.30.5.1131

Tily, H., Frank, M.C., & Jaeger, T.F. (2011). The learnability of constructed languages reflects typological patterns. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.*

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

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. DOI: 10.1016/j.lingua.2007.03.003

Williams, J.N., & Rebuschat, P. (2012). Statistical learning and syntax: What can be learned, and what difference does meaning make? In Rebuschat & Williams (Eds.) (pp. 237–264).

Witt, A., & Vinter, A. (2012). Artificial grammar learning in children: Abstraction of rules or sensitivity to perceptual features? *Psychological Research, 76*(1), 97–110. DOI: 10.1007/s00426-011-0328-5

Wright, R.L., & Burton, A.M. (1995). Implicit learning of an invariant: Just say no. *The Quarterly Journal of Experimental Psychology, 48A*(3), 783–796. DOI: 10.1080/14640749508401415

Wu, R., & Kirkham, N.Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology, 107*(2), 118–136. DOI: 10.1016/j.jecp.2010.04.014