

Introduction

Pattern perception and computational complexity: introduction to the special issue

W. Tecumseh Fitch^{1,*}, Angela D. Friederici² and Peter Hagoort^{3,4}

¹*Department of Cognitive Biology, University of Vienna, Althanstrasse 14, Vienna 1090, Austria*

²*Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany*

³*Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands*

⁴*Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

Research on pattern perception and rule learning, grounded in formal language theory (FLT) and using artificial grammar learning paradigms, has exploded in the last decade. This approach marries empirical research conducted by neuroscientists, psychologists and ethologists with the theory of computation and FLT, developed by mathematicians, linguists and computer scientists over the last century. Of particular current interest are comparative extensions of this work to non-human animals, and neuroscientific investigations using brain imaging techniques. We provide a short introduction to the history of these fields, and to some of the dominant hypotheses, to help contextualize these ongoing research programmes, and finally briefly introduce the papers in the current issue.

Keywords: comparative cognition; brain imaging; pattern perception; artificial grammar learning; formal language theory; theory of computation

1. INTRODUCTION: PROJECT GRAMMARAMA REVISITED

In 1957, George Miller [1] initiated a research programme at Harvard University to investigate rule-learning, in situations where participants are exposed to stimuli generated by rules, but are not told about these rules. The research programme was designed to understand how, given exposure to some finite subset of stimuli, a participant could ‘induce’ a set of rules that would allow them to recognize novel members of the broader set. The stimuli in question could be meaningless strings of letters, spoken syllables or other sounds, or structured images. Conceived broadly, the project was a seminal first attempt to understand how observers, exposed to a set of stimuli, could come up with a set of principles, patterns, rules or hypotheses that generalized over their observations. Such abstract principles, patterns, rules or hypotheses then allow the observer to recognize not just the previously seen stimuli, but a wide range of other stimuli consistent with them. Miller termed this approach ‘*pattern conception*’ (as opposed to ‘*pattern perception*’), because the abstract patterns in question were too abstract to be ‘truly perceptual’ [1].

Miller dubbed this research programme ‘Project Grammarama’, and he described its goals as follows: ‘Project Grammarama... is a program of laboratory experiments to investigate how people learn the grammatical rules underlying artificial languages’ [1]. Both ‘grammatical’ and ‘language’ are used technically here, in abstract ways that differ from normal usage. First, the ‘grammatical rules’ in such a system are drawn from the discipline of *formal language theory* (FLT), a body of mathematics central to the theory of computation and typically taught today in computer science curricula. A ‘grammar’ in FLT is any finite set of rules that, when combined, can generate a set of strings: the grammar is a ‘generative rule set’. Second, although the strings constituting ‘artificial languages’ in such studies are often termed ‘words’ or ‘sentences’, they are completely meaningless: they are simple abstract patterns of letters, syllables or images. Thus, in Miller’s and subsequent research programmes, terms such as grammar, language, word or sentence are used in an austere, abstract way, quite distant from their ordinary usage. In order to simplify and tame the problem of pattern conception enough to study it in the laboratory, this approach jettisons the entire context-dependent, interpretive dimension of meaning that is arguably the main point of real (natural) languages such as English, Dutch or Chinese.

Historically, Project Grammarama was seminal in several ways. In 1963, it became one of the first computer-based research programmes in experimental

* Author for correspondence (tecumseh.fitch@univie.ac.at).

One contribution of 13 to a Theme Issue ‘Pattern perception and computational complexity’.

psychology. One of its branches led to the ‘artificial grammar learning’ (AGL, sometimes also termed ‘synthetic grammar learning’) paradigm that has been a core tool for the study of implicit and explicit learning, starting with Arthur Reber [2]. Finally, the project introduced a generation of experimental psychologists to the arcane world of FLT: the branch of mathematics that concerns itself with finite algorithms, which generate potentially infinite sets of strings [3–5]. Nonetheless, as Miller humorously recounted in the last publication from the project in 1967, Project Grammmarama itself collapsed under its own weight, for several reasons. The first was that the number of possible questions to be addressed was vast, enough to keep a single laboratory busy for decades. The second problem was a worry that what was actually being explored was the cleverness of Harvard undergraduates, rather than the more general aspects of human rule conception that were Miller’s core interest. The third, and for Miller most crucial, was the meaninglessness of the strings: if one is interested in language acquisition, surely it would be more profitable to study the artificial languages with integrated meanings? For these reasons, Miller’s laboratory stopped research on AGL, and although many of its computerized tools and rule systems continued to be used, the grander scope of Miller’s original vision was, until recently, forgotten.

(a) *The ‘new wave’: Grammmarama for zoologists and neuroscientists*

In the last decade, there has been a major resurgence of interest in the Grammmarama approach, combining the mathematical precision of FLT with empirical research paradigms like AGL to probe ‘pattern conception’. There are two core advances that warrant this resurgence. The first is that the approach has been applied to a much wider range of study subjects, including pre-verbal infants and non-human animals. All of the early work was applied exclusively to adult humans, typically undergraduates, and it was not until the mid-1990s that the potential value of the AGL approach for studying infant cognition was realized [6], unleashing a flood of similar infant studies [7–10]. Later, in 2004, the first study applying this approach to non-human animals (‘animals’, hereafter) appeared [11], again spurring further studies on a variety of species [12–16].

The second major advance is in the domain of neuroscience, specifically brain imaging. Although simple grammars, of the type introduced by Reber, had been used with patient populations to explore implicit learning for years [17–19], the grammar itself was of only incidental interest in such studies: the focus was on the learning mode (implicit/explicit). The first paper where different formal classes of grammar were explored for their own sake was published in 2006 [20], and this has again led to a flood of further research [21–26] and considerable interest in the use of FLT in understanding neural computation [27].

The great interest that this new wave of research has elicited, across disciplines, is a testament to the enduring importance of the questions first raised by Miller in the 1960s. Unfortunately, this flood of papers has

also elicited a wave of critiques and counter-critiques, enough that an outsider becoming interested in this research programme might be discouraged from further exploration. The current issue is designed to avoid this unwanted outcome. It derives from an ERC-sponsored workshop held at the Max Planck Institute for Psycholinguistics in November 2010. The workshop participants encompassed a wide variety of approaches and included representatives of almost all the major laboratories involved in the new wave of AGL research. The participants shared the belief that it is worthwhile to investigate AGL from the viewpoint of more rigorous theory, and valuable to develop new study species and techniques, but they differed considerably on almost everything else. As a result, considerable, and ultimately fruitful, debate covered many detailed theoretical and empirical issues.

The goal of the current issue is to provide newcomers to this research area a concise and an authoritative overview into the literature and issues surrounding this controversial field and to outline the many outstanding questions it has opened. Two introductory tutorials are designed to help scientists to master the technical underpinnings of FLT and the theory of computation, allowing them to design their own experiments. The subsequent articles provide state-of-the-art overviews of the animal and neuroscientific literatures in AGL and present considerable new data, including several new species. It is our hope that these articles will help to convey some of the excitement and promise of this field, while also helping future researchers to avoid some of the pitfalls and confusion that arose in its first decade. The remainder of this short introduction will provide a concise history of its dual parent fields, FLT and AGL, and will close with a brief introduction to the articles in the current special issue.

2. FORMAL LANGUAGE THEORY AND ARTIFICIAL GRAMMAR LEARNING: A BRIEF HISTORICAL OVERVIEW

The theory of computation, of which FLT is a central component, has its roots in meta-mathematics, particularly the stunning achievements in the late 1930s of Alan Turing, Alonzo Church, Stephen Kleene and Emil Post, which defined the fundamental notions of ‘algorithm’ and ‘computability’ [28–30]. For historical reasons, the topic was often treated under the rubric ‘recursive function theory’ during the twentieth century [31], but this rather misleading term has slowly been replaced by ‘computability theory’ in the last decades [32,33]. Working separately, Turing and Church (with his student Kleene) developed three different mathematical formalisms to ground the idea of ‘computation’: the universal Turing machine, lambda-calculus and recursive function theory. It soon was proved that these all were equivalent, in that each of these formalisms could define exactly equivalent sets of functions [34]. A central outcome of this work was the widely accepted Church/Turing thesis, which equates the very word ‘computation’ with any algorithm that can be implemented on a Turing machine.

This early seminal work was concerned with defining the limits of mathematics, but the field took a more practical turn with Kleene's later work on regular (i.e. finite state) automata [35]. Such finite-state automata are the simplest systems capable of making 'infinite use of finite means', and thus the simplest that are of interest to mathematicians, computer scientists or linguists. Combining Turing machines and finite-state automata, Kleene's research provided upper and lower bounds on systems capable of computing infinite sets, with Turing machines representing the outer limit, and finite-state automata a limited subset. The notion that such models might be relevant to human natural languages (rather than mathematical formulae) was already apparent to Claude Shannon, the father of information theory. Shannon illustrated how very simple formal models could aid in quantifying the redundancy of English [36].

FLT took a stronger linguistic turn after the contributions of the young Noam Chomsky and colleagues [3,37]. Chomsky first provided a compelling argument that finite-state approximations of the sort investigated by Shannon were inadequate to capture the syntax of English (or, by inference, any other human language). Second, he proposed two new systems of computation, intermediate in power between finite-state automata and full-blown Turing machines. These new levels were termed 'context-free' and 'context-sensitive' systems. Together, these four classes of systems can be arranged in a hierarchy of increasing computational power (traditionally termed the *Chomsky hierarchy*) and discussed in later articles (figure 1).

(a) *The supra-regular hypothesis*

Psychologist George Miller was apparently the first experimentalist to see the potential empirical value of this rather daunting branch of theoretical mathematics, and working together with Chomsky, he helped package and present it to a wider audience [39–41]. Miller's first Grammmarama project was a study on participants' capacity to memorize the strings generated by a finite-state grammar [42]. Later, convinced by Chomsky's arguments about the inadequacy of finite-state automata as models for linguistic syntax, Miller *et al.* [1] also began experimenting with various rule sets at the context-free level and found not only that people are capable of learning such higher-level systems, but, in many cases, they actually find these computationally more challenging systems *easier* to master. This work ultimately led to an intriguing hypothesis about the nature of human rule learning, which we will term Miller's *supra-regular hypothesis*:

'constituent structure languages are more natural, easier to cope with, than regular languages... The hierarchical structure of strings generated by constituent-structure grammars is characteristic of much other behavior that is sequentially organized; it seems plausible that it would be easier for people than would the left-to-right organization characteristic of strings generated by regular grammars' [1]

Miller uses the term 'constituent structure' here to refer to what today are called 'context-free' and other supra-regular systems. The key insight underlying

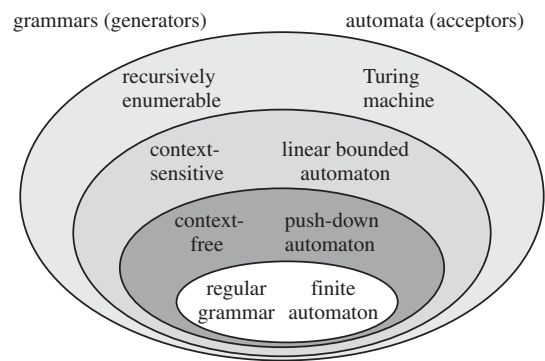


Figure 1. The Traditional Chomsky hierarchy outlines four classes of 'grammars', or sets of rules, with increasing power, along with a matched set of 'formal machines' or 'automata' that exactly correspond to the adjacent grammars. This formal language hierarchy has been extended considerably; see Jäger & Rogers [38].

this hypothesis about human cognition is that the kind of sequential, left-to-right structures generated by regular (i.e. finite state) grammars is inadequate to capture our typical or 'natural' cognitive approach to learning rules about some set of patterned strings. By Miller's psychological hypothesis, which derives directly from Chomsky's formal argument, some additional cognitive resources, more powerful than those offered by regular grammars, are required for human language and human pattern conception more generally.

Among linguists, psycholinguists and computer scientists today, the supra-regular hypothesis is nearly universally accepted. Supra-regular grammars, and context-free grammars in particular, have become standard and indispensable components of computer science [43] and computational linguistics [44]. Today, the supra-regular hypothesis is seen as axiomatic in computer science: for example, no one today would attempt to use anything less powerful than a context-free grammar to define a computer programming language [45], and probabilistic context-free grammars are widely used in computational models of human grammar induction [46]. Similarly, at least since the seminal review by Levelt [5], psycholinguists also typically assume that supra-regular grammars are needed to adequately model human syntax.

An intensive debate about whether context-free grammars by themselves were adequate for the syntax of natural languages [47] was eventually resolved negatively by Huybregts [48]. Computational linguists widely agree today that supra-regular grammars called 'mildly context-sensitive grammars', a new layer of the formal language hierarchy just beyond the power of context-free grammars, are required to deal with all of the syntactic phenomena of natural language. However, the vast majority of syntactic constructions of most languages can still be captured by context-free grammars. Interestingly, it has been shown that multiple different syntactic formalisms, developed independently for different purposes by different scholars, converge at this mildly context-sensitive (MCS) level [49,50]. This example of intellectual 'convergent evolution',

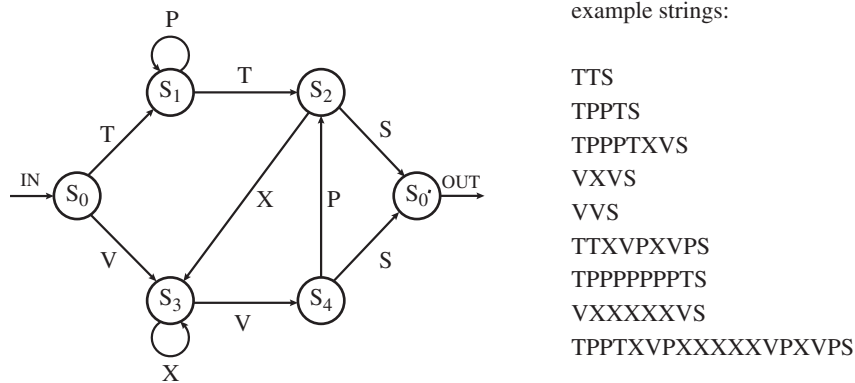


Figure 2. The Reber grammar. The vast majority of research in artificial grammar learning has used this moderately complex regular grammar of Reber [2], illustrated here by its corresponding finite-state machine (after Reber [2]).

reminiscent of the earlier convergence of Church, Turing and Kleene's formalisms, suggests that this issue has finally been resolved: human languages are supra-regular at the MCS level.

Thus, among linguists and computer scientists, it has become a truism that natural language requires supra-regular resources, which are thus presumed to be present in some form in the human mind and implemented by human brains: this is not an issue debated in the recent literature. It is thus a peculiar historical fact that, until very recently, neither neuroscientists nor experimental and animal psychologists have shown any interest in this issue.

(b) *Experimental psychology: artificial grammar learning and implicit learning*

Unfortunately, once Miller ended the Grammmarama project, virtually all experimental work in AGL returned to studies of regular finite-state grammars, which were used simply as tools to investigate learning and memory [51]. Regular grammars were used mainly because these were easy to work with, and produced stringsets adequate for the study of implicit memory processes [2,52]: the grammars in such studies were arbitrary and of little or no interest in themselves.

In a typical AGL study, a grammar such as that illustrated in figure 2 is used to generate a set of meaningless letter strings. Individual strings are briefly presented to participants, who are required to write or type them as accurately as possible. During this 'training' phase, the subjects are not told that the strings are generated by rules or follow any pattern. Later, during the 'test' phase, they are told that the previous strings followed an underlying pattern, and they are asked to judge whether novel strings fit, or violate, that pattern. A typical, and very robust, finding is that participants are able to discriminate 'grammatical' from 'ungrammatical' stimuli at a level far above chance, usually between 70 and 80 per cent correct. Nonetheless, they are typically unable to explicitly state the rules they use to achieve this task, and the knowledge that they can state is fragmentary or even incorrect (reviewed in Reber [52]).

This research topic ballooned into a now extensive research interest in implicit versus explicit learning and memory, which has close ties to the question of exemplar- versus rule-based learning [51]. After some debate in the 1990s concerning the degree to which the knowledge obtained by mere exposure in AGL experiments is truly 'implicit' and unavailable to introspection [53–55], it is now clear that most participants can perform well on such tasks with little conscious understanding of the rules, and that they do so by forming some abstract rules and not just relying on analogy [56] or bi-gram and tri-gram chunks [54]. This demonstration was made most convincingly by Barbara Knowlton and Larry Squire [17,57], who showed not only that normal participants could successfully perform in the absence of analogical and chunk cues, but also that amnesic patients could do the same. This seminal work led to the first wave of neuroscientific (mostly patient-based) research in the AGL tradition, focused on the notion of two independent memory systems for implicit or 'procedural' memory (a neo-striatal system) and explicit or 'declarative' memory (based in the medial temporal lobe) [18,19,58].

Two important methodological innovations from this research tradition are worth noting. First, it is possible to control for the possibility that bi- or tri-grams are the sole basis for learning in an AGL experiment by explicitly balancing the novel strings for a measure called 'associative chunk strength' [17]. This measure essentially tabulates and averages the probability of each possible bi- or tri-gram in the training data, and thus allows a computation of the probability of novel strings in the test data. By equalizing this probability between correct and incorrect strings (which can be done by putting correct chunks in the incorrect locations), it is possible to rule out a simple associative memory explanation for AGL performance [59]. A second innovation is that it is possible to avoid *ever* telling participants that there is any rule system involved in the study. This is most easily done by giving the subjects a preference task in the test phase (rather than asking them to make grammaticality judgements, a categorization task) [26,60]. Such experiments have demonstrated that such preferences match the grammatical strings, though slightly

less robustly than standard explicit categorization. This new approach is quite relevant for studies of animal or infant subjects, who obviously cannot be asked to make explicit grammaticality judgements.

In summary, despite a significant and ongoing body of research in AGL within psychology and neuroscience, the grammars themselves have been of little or no interest (and indeed, the vast majority of studies use the same regular grammar as the original Reber studies) [61]. As a result, some of the fundamental questions that drove the development of regular grammars or that fuelled Miller's research programme have been forgotten, and experimental research focused almost entirely on questions of implicit learning.

(c) *How biologically and neurally distinctive is supra-regularity?*

This changed quite rapidly in the last decade, when Fitch & Hauser [11] tested a comparative variant of Miller's hypothesis in monkeys, the core question being to what extent the supra-regular syntactic capabilities possessed by humans are also available to other animals. They used a 'mere exposure' AGL-like paradigm with cotton-top tamarin monkeys, *Saguinus oedipus*, as a test species. Their results showed that these monkeys were able to master a very simple regular grammar, but failed to distinguish grammatical from ungrammatical strings when they were generated by a simple supra-regular grammar. Thus, for this primate species, they answered this question in the negative, leading them to state the *supra-regular distinctiveness hypothesis*: that humans are unusual (or perhaps unique) among animals in possessing supra-regular processing power.

Shortly thereafter, Gentner *et al.* [13] tested this hypothesis using very similar materials, but in a drawn-out operant training paradigm, to give a positive result for starlings, *Sturnus vulgaris*. A host of further studies with other species have followed (see the review by ten Cate & Okanoya [62]). Summing up these data, it currently seems plausible to suppose that most animal species do not display the tendency Miller observed in humans, to find a supra-regular grammar 'more natural and easier to cope with'. The only two animal species to show any evidence of supra-regular abilities require intensive training to do so, and even then the involved researchers differ in their interpretation of this evidence. Though this research area remains a hotbed of controversy, a plausible working hypothesis, at present, is that humans are biologically quite unusual in our propensity to infer supra-regular structure from a set of strings, with no encouragement or training.

This initial comparative data also led to a renewed interest in the supra-regular hypothesis by experimental psychologists and neuroscientists. By a related hypothesis, which we might term the *supra-regular neural distinctiveness hypothesis*, Friederici and co-workers suggested that the human capacity for supra-regular grammars may be implemented in brain regions separate from those involved in simple sequencing at the sub-regular level. These hypotheses, and the controversy surrounding them, are explored by many of

the papers in this issue, and it is fair to say that no consensus on this hypothesis is currently available.

3. THE CURRENT ISSUE: OVERVIEW

The current issue is designed to encourage further detailed exploration of rule learning and 'pattern conception' in a wide variety of species, studied using a wide variety of neuroscientific techniques and approaches. The issue has three main goals. The first is to provide a concise introduction to the background information required to understand the existing literature and design experiments that go beyond the current state of the art. This necessitates an understanding of AGL and related paradigms, and more dauntingly of the technical apparatus of FLT that provides the theoretical framework for this literature. This goal will first be approached informally in the introduction by Fitch & Friederici [63] and then with more formal rigour by Jäger & Rogers [38]. Together, these two papers provide a basic introduction to FLT that will be understandable to practising biologists, neuroscientists or psychologists, and should help in the future to avoid unproductive debates about terminology or theory.

The second goal is to concisely review the (mostly recent) literature on AGL in animals [62] and brain imaging [63–65]. One paper further offers a first glimpse at work combining these two approaches, using brain imaging of monkeys [66]. Two empirical papers, which directly compare human and animal performance, introduce novel techniques involving 'visual grammar' learning and provide data supporting the supra-regular distinctiveness hypothesis for pigeons and parrots [67,68].

The third and final goal is to survey new empirical approaches and outstanding theoretical issues. One line of debate has concerned the degree to which FLT, with its focus on potentially infinite sets of strings, is relevant to ours' or animals' brains. This issue is explored in detail by Fitch & Friederici [63] and Petersson & Hagoort [69]. Another concerns the degree to which formal 'competence' models can be reconciled with performance models [70]. Finally, the fraught issue of what, if anything, this whole line of research has to do with 'recursion' is covered by the contributions from Martins [71] and from Poletiek & Lai [72].

In summary, we aim to provide in this issue a concise introduction to the promise, and potential pitfalls, of research at the intersection of AGL and FLT. Despite their differences, the authors of this special issue share a conviction that this approach is interesting and valuable, and share a desire to bypass fruitless debates about terminology or philosophy to get to the interesting hypotheses and issues that will drive this research programme forward.

Several important limitations of the AGL + FLT research programme remain with respect to human studies (for example, for research in psycholinguistics). These include the implicit nature of learning, the lack of meaning, and the focus on strings rather than structures. Nonetheless, in the modern context of animal and brain research, these limitations can often be

seen as virtues. Animals or non-verbal infants cannot explicitly tell us what rules they have induced, so the implicit nature of AGL is well-suited for these new classes of participants. The lack of meaning, while a limitation for approaches oriented towards human language, makes the FLT approach a good fit for abstract visual patterns, for music, or for complex animal signals such as bird song.

In human brain research, AGL + FLT can help to isolate the neural infrastructure for syntax and sequence processing from the need for to semantic interpretation. In normal language, processing syntax and semantics are intricately related and temporally overlapping. It is thus far from trivial to distinguish the neural infrastructure of syntax from the mapping of grammatical roles onto thematic roles and further semantic processing. Here, AGL can be useful due to the absence of any semantic consequences of syntactic decision-making.

Finally, in AGL paradigms, we typically have access to simple yes/no decisions concerning strings. In contrast, the ultimate goal in linguistics is clearly to understand syntactic structures rather than flat strings [73]. Nonetheless, for non-verbal participants such as infants and animals, potentially with quite short attention spans, it is far easier to design experiments involving acceptance or rejection of strings, than to attempt to determine the structures they might ultimately attribute to such strings. Whether these experiments necessarily inform us about the processing of underlying syntactic structures or the processing of surface dependencies needs to be worked out in future studies.

In conclusion, the research discussed in this issue breathes new life into a set of issues that were raised, but never resolved, by Miller 60 years ago and offers several productive new routes to overcoming problems that Miller faced in Project Grammmarama. In particular, FLT offers a mathematical ‘ground truth’, as well as a well-developed terminology and descriptive apparatus, that offers a valuable complement to the empirical research methodology of AGL. The articles in this issue nicely illustrate the value of this new combination in sharpening hypotheses and analysing results, especially when applied to new species, or as employed in brain imaging paradigms.

REFERENCES

- 1 Miller, G. A. 1967 Project Grammmarama. In *Psychology of communication* (ed. G. A. Miller). New York, NY: Basic Books.
- 2 Reber, A. S. 1967 Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* **6**, 855–863. (doi:10.1016/S0022-5371(67)80149-X)
- 3 Chomsky, N. 1956 Three models for the description of language. *IRE Trans. Inform. Theory* **2**, 113–124. (doi:10.1109/TIT.1956.1056813)
- 4 Minsky, M. L. 1967 *Computation: finite and infinite machines*. Englewood Cliffs, NJ: Prentice-Hall.
- 5 Levelt, W. J. M. 1974 *Formal grammars in linguistics and psycholinguistics: volume 1: an introduction to the theory of formal languages and automata, volume 2: applications in linguistic theory, volume 3: psycholinguistic applications*. The Hague, The Netherlands: Mouton.
- 6 Saffran, J. R., Aslin, R. N. & Newport, E. L. 1996 Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)
- 7 Saffran, J. R., Johnson, E., Aslin, R. N. & Newport, E. L. 1999 Statistical learning of tone sequences by human infants and adults. *Cognition* **70**, 227–252. (doi:10.1016/S0010-0277(98)00075-4)
- 8 Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. 1999 Rule learning by seven-month-old infants. *Science* **283**, 77–80. (doi:10.1126/science.283.5398.77)
- 9 Gómez, R. L. & Gerken, L. 1999 Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* **70**, 109–135. (doi:10.1016/S0010-0277(99)00003-7)
- 10 Gervain, J., Nespors, M., Mazuka, R., Horle, R. & Mehler, J. 2008 Bootstrapping word order in prelexical infants: a Japanese–Italian cross-linguistic study. *Cogn. Psychol.* **57**, 56–74. (doi:10.1016/j.cogpsych.2007.12.001)
- 11 Fitch, W. T. & Hauser, M. D. 2004 Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380. (doi:10.1126/science.1089401)
- 12 Abe, K. & Watanabe, D. 2011 Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* **14**, 1067–1074. (doi:10.1038/nn.2869)
- 13 Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. 2006 Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207. (doi:10.1038/nature04675)
- 14 Herbranson, W. T. & Shimp, C. P. 2008 Artificial grammar learning in pigeons. *Learn. Behav.* **36**, 116–137. (doi:10.3758/LB.36.2.116)
- 15 Murphy, R. A., Mondragón, E. & Murphy, V. A. 2008 Rule learning by rats. *Science* **319**, 1849–1851. (doi:10.1126/science.1151564)
- 16 van Heijningen, C. A. A., de Vissera, J., Zuidema, W. & ten Cate, C. 2009 Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl Acad. Sci. USA* **106**, 20 538–20 543. (doi:10.1073/pnas.0908113106)
- 17 Knowlton, B. J. & Squire, L. R. 1996 Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 169–181. (doi:10.1037/0278-7393.22.1.169)
- 18 Squire, L. R., Knowlton, B. J. & Musen, G. 1993 The structure and organization of memory. *Annu. Rev. Psychol.* **44**, 453–495. (doi:10.1146/annurev.ps.44.020193.002321)
- 19 Reber, P. J. & Squire, L. R. 1999 Intact learning of artificial grammars and intact category learning by patients with Parkinson’s disease. *Behav. Neurosci.* **113**, 235–242. (doi:10.1037/0735-7044.113.2.235)
- 20 Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I. & Anwander, A. 2006 The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl Acad. Sci. USA* **103**, 2458–2463. (doi:10.1073/pnas.0509389103)
- 21 Bahlmann, J., Gunter, T. C. & Friederici, A. D. 2006 Hierarchical and linear sequence processing: an electrophysiological exploration of two different grammar types. *J. Cogn. Neurosci.* **18**, 1829–1842. (doi:10.1162/jocn.2006.18.11.1829)
- 22 Bahlmann, J., Schubotz, R. I. & Friederici, A. D. 2008 Hierarchical artificial grammar processing engages Broca’s area. *Neuroimage* **42**, 525–534. (doi:10.1016/j.neuroimage.2008.04.249)
- 23 Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D. & Friederici, A. D. 2009 Neural circuits of hierarchical

- visuo-spatial sequence processing. *Brain Res.* **1298**, 161–170. (doi:10.1016/j.brainres.2009.08.017)
- 24 Makuuchi, M., Bahlmann, J., Anwander, A. & Friederici, A. D. 2009 Segregating the core computational faculty of human language from working memory. *Proc. Natl Acad. Sci. USA* **106**, 8362–8367. (doi:10.1073/pnas.0810928106)
- 25 Udden, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P. & Petersson, K. M. 2008 The inferior frontal cortex in artificial syntax processing: an rTMS study. *Brain Res.* **1224**, 68–79. (doi:10.1016/j.brainres.2008.05.070)
- 26 Udden, J., Ingvar, M., Hagoort, P. & Petersson, K. M. 2012 Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model. *Cogn. Sci.* (doi:10.1111/j.1551-6709.2012.01235.x)
- 27 Pulvermüller, F. 2010 Brain embodiment of syntax and grammar: discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain Lang.* **112**, 167–179. (doi:10.1016/j.bandl.2009.08.002)
- 28 Turing, A. M. 1937 On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **42**, 230–265. (doi:10.1112/plms/s2-42.1.230)
- 29 Post, E. L. 1944 Recursively enumerable sets of positive integers and their decision problems. *Bull. Am. Math. Soc.* **50**, 284–316. (doi:10.1090/S0002-9904-1944-08111-1)
- 30 Kleene, S. C. 1936 Lambda-definability and recursiveness. *Duke Math. J.* **2**, 340–353. (doi:10.1215/S0012-7094-36-00227-2)
- 31 Cutland, N. 1980 *Computability: an introduction to recursive function theory*. Cambridge, UK: Cambridge University Press.
- 32 Epstein, R. L. & Carnielli, W. A. 2000 *Computability: computable functions, logic, and the foundations of mathematics*, 2nd edn. London, UK: Wadsworth/Thomson Learning.
- 33 Soare, R. I. 1996 Computability and recursion. *Bull. Symbol. Logic* **2**, 284–321. (doi:10.2307/420992)
- 34 Davis, M. 1958 *Computability and unsolvability*. New York, NY: McGraw-Hill.
- 35 Kleene, S. C. 1956 Representation of events in nerve nets and finite automata. In *Automata studies* (eds C. E. Shannon & J. J. McCarthy), pp. 3–40. Princeton, NJ: Princeton University Press.
- 36 Shannon, C. E. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379, 623–656.
- 37 Chomsky, N. 1957 *Syntactic structures*. The Hague, The Netherlands: Mouton.
- 38 Jäger, G. & Rogers, J. 2012 Formal language theory: refining the Chomsky hierarchy. *Phil. Trans. R. Soc. B* **367**, 1956–1970. (doi:10.1098/rstb.2012.0077)
- 39 Chomsky, N. & Miller, G. A. 1958 Finite state languages. *Inform. Control* **1**, 91–112. (doi:10.1016/S0019-9958(58)90082-2)
- 40 Chomsky, N. & Miller, G. A. 1963 Introduction to the formal analysis of natural languages. In *Handbook of mathematical psychology* (eds R. D. Luce, R. R. Bush & E. Galanter), pp. 269–322. New York, NY: John Wiley & Sons.
- 41 Miller, G. A. & Chomsky, N. 1963 Finitary models of language users. In *Handbook of mathematical psychology* (eds R. D. Luce, R. R. Bush & E. Galanter), pp. 419–492. New York, NY: John Wiley & Sons.
- 42 Miller, G. A. 1958 Free recall of redundant strings of letters. *J. Exp. Psychol.* **57**, 485–491. (doi:10.1037/h0044933)
- 43 Hopcroft, J. E., Motwani, R. & Ullman, J. D. 2000 *Introduction to automata theory, languages and computation*, 2nd edn. Reading, MA: Addison-Wesley.
- 44 Jurafsky, D. & Martin, J. H. 2000 *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.
- 45 Gersting, J. L. 1999 *Mathematical structures for computer science*, 4th edn. New York: W H Freeman.
- 46 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)
- 47 Pullum, G. K. & Gazdar, G. 1982 Natural languages and context-free languages. *Linguist. Philos.* **4**, 471–504. (doi:10.1007/BF00360802)
- 48 Huybregts, R. 1985 The weak inadequacy of CFPSGs. In *Van Periferie naar Kern* (eds G. de Haan, M. Trommelen & W. Zonneveld), pp. 81–99. Dordrecht, The Netherlands: Foris Publications.
- 49 Stabler, E. P. 2004 Varieties of crossing dependencies: structure dependence and mild context sensitivity. *Cogn. Sci.* **28**, 699–720. (doi:10.1207/s15516709cog2805_4)
- 50 Vijay-Shanker, K. & Weir, D. J. 1994 The equivalence of four extensions of context-free grammars. *Math. Syst. Theory* **27**, 511–546. (doi:10.1007/BF01191624)
- 51 Pothos, E. M. 2007 Theories of artificial grammar learning. *Psychol. Bull.* **133**, 227–244. (doi:10.1037/0033-2909.133.2.227)
- 52 Reber, A. S. 1989 Implicit learning and tacit knowledge. *J. Exp. Psychol. Gen.* **118**, 219–235. (doi:10.1037/0096-3445.118.3.219)
- 53 Mathews, R. C. 1990 Abstractness of implicit grammar knowledge: comments on Perruchet and Pacteau’s analysis of synthetic grammar learning. *J. Exp. Psychol. Gen.* **119**, 412–416. (doi:10.1037/0096-3445.119.4.412)
- 54 Perruchet, P. & Pacteau, C. 1990 Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *J. Exp. Psychol. Gen.* **119**, 264–275. (doi:10.1037/0096-3445.119.3.264)
- 55 Reber, A. S. 1990 On the primacy of the implicit: a comment on Perruchet and Pacteau. *J. Exp. Psychol. Gen.* **119**, 340–342. (doi:10.1037/0096-3445.119.3.340)
- 56 Brooks, L. R. & Vokey, J. R. 1991 Abstract analogies and abstracted grammars: comments on Reber (1989) and Mathews et al. (1989). *J. Exp. Psychol. Gen.* **120**, 316–323. (doi:10.1037/0096-3445.120.3.316)
- 57 Knowlton, B. J. & Squire, L. R. 1994 The information acquired during artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **20**, 79–91. (doi:10.1037/0278-7393.20.1.79)
- 58 Knowlton, B. J., Mangels, J. A. & Squire, L. R. 1996 A neostriatal habit learning system in humans. *Science* **273**, 1399–1402. (doi:10.1126/science.273.5280.1399)
- 59 Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M. & Petersson, K. M. 2006 Neural correlates of artificial syntactic structure classification. *Neuroimage* **32**, 956–967. (doi:10.1016/j.neuroimage.2006.03.057)
- 60 Forkstam, C., Elwér, A., Ingvar, M. & Petersson, K. M. 2008 Instruction effects in implicit artificial grammar learning: a preference for grammaticality. *Brain Res.* **1221**, 80–92. (doi:10.1016/j.brainres.2008.05.005)
- 61 Forkstam, C. & Petersson, K. M. 2005 Towards an explicit account of implicit learning. *Curr. Opin. Neurol.* **18**, 435–441. (doi:10.1097/01.wco.0000171951.82995.c4)
- 62 ten Cate, C. & Okanoya, K. 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Phil. Trans. R. Soc. B* **367**, 1984–1994. (doi:10.1098/rstb.2012.0055)
- 63 Fitch, W. T. & Friederici, A. D. 2012 Artificial grammar learning meets formal language theory: an overview. *Phil. Trans. R. Soc. B* **367**, 1933–1955. (doi:10.1098/rstb.2012.0103)
- 64 Makuuchi, M., Bahlmann, J. & Friederici, A. D. 2012 An approach to separating the levels of hierarchical structure

- building in language and mathematics. *Phil. Trans. R. Soc. B* **367**, 2033–2045. (doi:10.1098/rstb.2012.0095)
- 65 Uddén, J. & Bahlmann, J. 2012 A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Phil. Trans. R. Soc. B* **367**, 2023–2032. (doi:10.1098/rstb.2012.0009)
- 66 Petkov, C. I. & Wilson, B. 2012 On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Phil. Trans. R. Soc. B* **367**, 2077–2088. (doi:10.1098/rstb.2012.0073)
- 67 Stobbe, N., Westphal-Fitch, G., Aust, U. & Fitch, W. T. 2012 Visual artificial grammar learning: comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). *Phil. Trans. R. Soc. B* **367**, 1995–2006. (doi:10.1098/rstb.2012.0096)
- 68 Westphal-Fitch, G., Huber, L., Gomez, J. C. & Fitch, W. T. 2012 Production and perception rules underlying visual patterns: effects of symmetry and hierarchy. *Phil. Trans. R. Soc. B* **367**, 2007–2022. (doi:10.1098/rstb.2012.0098)
- 69 Petersson, K. M. & Hagoort, P. 2012 The neurobiology of syntax: beyond string sets. *Phil. Trans. R. Soc. B* **367**, 1971–1983. (doi:10.1098/rstb.2012.0101)
- 70 de Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P. & Christiansen, M. H. 2012 Processing multiple non-adjacent dependencies: evidence from sequence learning. *Phil. Trans. R. Soc. B* **367**, 2065–2076. (doi:10.1098/rstb.2011.0414)
- 71 Martins, M. D. 2012 Distinctive signatures of recursion. *Phil. Trans. R. Soc. B* **367**, 2055–2064. (doi:10.1098/rstb.2012.0097)
- 72 Poletiek, F. H. & Lai, J. 2012 How semantic biases in simple adjacencies affect learning a complex structure with non-adjacencies in AGL: a statistical account. *Phil. Trans. R. Soc. B* **367**, 2046–2054. (doi:10.1098/rstb.2012.0100)
- 73 Chomsky, N. 1990 On formalization and formal linguistics. *Nat. Lang. Linguist. Theory* **8**, 143–147.