C O P *i* L CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **14(2)**, Article 5: pp. 165–175, 2022 | ISSN: 2050-5949

Hierarchical Syntax and Domain-Generality: A Review

Dag-Tore Nordbø
Kristiansen
University of Cambridge

1 Introduction

This literature review paper will investigate whether the hierarchical fundamentals of generative syntax such as non-adjacent dependencies (NADs) and recursion can be accounted for solely by domain-general processes.

In reviewing this, I will start by briefly describing the fundamentals of generative grammar that pertain to NADs and recursion. The bulk of this paper will focus on empirical approaches to investigating the possibility of domain-general learning and processing of these structures. Firstly, I will look at evidence from artificial language (AL) experiments on statistical learning. Subsequently, I will discuss connectionist models such as neural networks, and how they may provide more controlled environments for investigating this question. In tandem with this, I will discuss a model for learning that was introduced in this paradigm: incremental learning. Following this, I will review some of the neurocognitive research on syntactic processing. In my conclusion, I will attempt to synthesise these findings from different areas, and suggest some possible avenues for further research.

2 Recursion and NADs in Generative Syntax

Generative syntax arose as an answer to the inherent and infinite productivity of language, leading to the assumption that the environment cannot provide enough stimulus for reinforcement of observed regularities in grammar (Chomsky 1959a). There are thus features of language that are proposed to be innate, and these have by some been argued to be a fact of human biology and natural selection (Pinker & Bloom 1990). Generative grammar tries to find these in theory, by providing a set of formally defined laws derived from a finite set of examples, which should be able to make successful predictions for the language they are applied to (Chomsky 1956: 113).

Furthermore, Chomsky argues that the productivity and hierarchical structure of natural language goes beyond what is permissible in a finite state grammar, as the latter disallow potentially infinite embeddings (Chomsky 1956: 115-116; Chomsky 1959b: 137-141). In later incarnations of generative syntax, this recursivity was essentialised as the fundamental operation of Merge (Chomsky 1995), which is in essence the faculty of language in the narrow sense (FLN), and thus where the constraints of UG apply, while other cognitive processes and the environment provide the other functions and constraints upon language processing (Hauser, Chomsky & Fitch 2002).

This universalisation of recursion begs the question: are recursive structures typologically universal? Many polysynthetic languages show limited embedding; they are often morphologically complex, but syntactically simple (Evans & Levinson 2009: 442-443). The most radical and controversial refutation of universal recursion came from Everett's (2005) study of Pirahã, an Amazonian language that does not seem to allow any embedded or subordinated clauses, nor recursive possessive phrases (pp. 628-631). Even in languages which allow it, such as many European languages, recursion is not unbounded. This is most clearly illustrated by higher-order centre-embedded structures such as relative clauses, which are rarely attested beyond one recursive level (Christiansen & Chater 2015: 3-4). Thus, it is not an infinite computational mechanism in practice.

There are other syntactic structures that involve non-linear relations, often described in theory by a form of binding (Chomsky 1981, Reinhart 1976), namely non-adjacent dependencies. From morphosyntactic phenomena such as case and agreement, to the binding of anaphora and reflexives: all involve a hierarchical relation that may transcend linear adjacency, and do not seem to find an explanation outside of grammar (Tallerman, Newmeyer, Bickerton, Bouchard, Kaan & Rizzi 2009: 140-145). Thus, while both recursion and non-adjacent dependencies are not necessarily infinite in practice, they are to different extents productive and prevalent in natural languages, and proponents of domain-general acquisition and processing of syntax must be able to find a reasonable mechanism by which these structures can be acquired without any innate language-specific machinery. These approaches will be discussed in the following sections.

3 Statistical Learning in AL Experiments

Already in the 60s, Reber (1967) proposed that an artificial language derived from a finite state grammar, which would contain statistical distribution information about its transitions, could be used to model language learning as a process of differentiation over stimulus-inherent information. In the 90s, these AL experiments had their breakthrough. Increased transitional probability between syllabic sequences was found to allow infants to distinguish word boundaries (Saffran, Aslin & Newport 1996), or to differentiate between word-like units (Saffran 2001, 2002). Crucially, it was found that the same learning mechanism could be recruited for domains beyond the linguistic, with tonal sequences in an adult participant experiment (Saffran, Johnson, Aslin & Newport 1999), or visual colour-shape patterns in another experiment with infants (Kirkham, Slemmer & Johnson 2002). The same sensitivity to transitional probabilities was discovered in other species, such as tamarin monkeys (Hauser, Newport & Aslin 2001).

Nevertheless, these experiments are all uncovering statistical relations between adjacent and linear sequences. One of the strengths of the generative framework is its hierarchical and recursive nature which transcends linear sequences. Can statistical learning be a mechanism of learning for these more complex structures?

NADs have been shown to be acquired in AL experiments with both children and adults, but this was limited to conditions where there is significant variation between the intervening elements (Gómez 2002). This is suggested to imply that learners default to focusing on transitional probabilities between adjacent elements, and will only switch to tracking non-adjacent elements provided that the adjacent elements are sufficiently variable. Previous exposure to NADs has been shown to help participants learn ALs which show less variance in the intervening elements (Zettersten, Potter & Saffran 2020). This could also allow for correct processing of NADs that are much less frequent in language, and thus be part of a potential domain-general solution to the poverty of the stimulus problem. These were all experiments using only nonce linguistic data, thus generalisability to non-linguistic domains is not proven.

In the phonological (Newport & Aslin 2004), and also non-linguistic (musical) (Gebhart, Newport & Aslin 2009) domains, statistical learning seems to be constrained as an acquisition mechanism for NADs, unless dependencies followed from perceptual similarity cues such as consonants and vowels. Differences in human and tamarin monkeys in which perceptual categories NADs could be acquired were also found (Newport, Hauser, Spaepen & Aslin 2004). There thus seems to be limits to the strength and generalisability of NAD statistical learning, more so than with adjacent elements.

Recursion has also been investigated. One example is Fitch & Hauser's (2004) study, which shows that humans are shown to learn a centre-embedded recursive structure such as an AnBn grammar, which is a structure that goes beyond the finite state grammar design of Reber (1967). Conversely, tamarin monkeys are not able to discriminate sentences based on familiarisation with this grammar, only for a simpler adjacent-dependency sequential grammar (Fitch & Hauser 2004). Critics have pointed out that this study's experimental design did not guarantee that the human participants actually learned a centre-embedded structure with mapping between A to B at each level, and a replication study suggested that they might have discriminated based on the acoustic differences between A and B syllables in the experiment, and not the supposedly centre-embedded structures (Perruchet & Rey 2005). The evidence for instantaneous learning of hierarchical structures by humans, whether through use of a language faculty or not, thus seems to be lacking.

There has also been criticism of the AL methodological approach itself. The AL design is limited, and has often employed a finite state grammar design, which is a simpler system than Chomskian phrase structure grammar (Tao & Williams 2018: 1001-1002). It has also been shown that AL learning and incidental SLA diverge markedly in the cognitive domain (Robinson 2005). Given these limitations, these experiments are not enough to prove that domain-general mechanisms such as transitional probability are powerful enough to capture the full complexity of natural language (Yang 2004). Yang instead suggests that UG initially constrains the learner to be attentive to certain statistical patterns over others.

The suggested combination of statistical learning and domain-specific innate constraints could be formulated in a probabilistic system (Yang 2004: 452-454). An example of this is found in probabilistic harmonic grammar, a Bayesian modification of Optimality Theory. There, the previously forbidden constraints upon parameter setting in UG are penalised, but not outlawed; the more acceptable ones are given a higher probability to occur, all given the prior of the innate language faculty. This system can thus account for some of the strong tendencies in, for example, word order as seen in typology, and shows similarly successful AL learning results (Culbertson, Smolensky & Wilson 2013).

It is clear that language-like patterns, even those as complex as non-adjacent dependencies, are possible to acquire through statistical learning, although invariant elements or specific perceptual cues are needed. Furthermore, it does not seem to be as generalisable to other mammals. As the languages used are simple, both in form and structure, the findings' applicability to natural language remains opaque. An experimental 'learner' that can fully rule out any innate linguistic knowledge would thus be useful.

4 Connectionist Models and Incremental Learning

An approach that allows for an artificial 'learner' of this kind can be found in neural networks. Neural networks consist of units analogous to neurons, with input and output units at each end, and hidden units in between. They are linked together, and the strength of these connections can be modified by the learning algorithm like adjusting weights, providing information that is integrated in parallel by each neuron (Abdi 1994). If the strength of the connections is randomly set at first, then it is possible to verify that learning is not based on any domain-specific system, but defined by the initially domain-general statistical learning algorithm. Neural networks are particularly sensitive to early data, as that is when the initial constraints for the hypothesis space are set (Elman 1993: 85-95).

Elman (1993) investigated NAD and recursion acquisition using recurrent neural networks (RNNs), which include a set of context units that feed the previous hidden unit activations back into the network together with the new input (Elman 1991: 95-96). RNNs struggled to learn these more complex structures, but they were considerably more successful when incremental learning was implemented, which entailed starting with simple sentences, and progressing to higher levels of embedding later. This was thought by Elman to not be as representative of the natural learning environment, as he suggested children are still exposed to mostly adult language during acquisition. He hypothesised that incremental learning could instead be realised by restricting the memory of the network at first, then gradually increasing it, thus simulating internal constraints upon learning. Similarly successful results were obtained for this framework (Elman 1993).

Incremental learning might show success in neural networks due to statistical information being easier to encode when the (perceived) input is simpler, and this bare structure can then be generalised to more complex structure. An equivalent mechanism could be available for children, and might explain their quick acquisition of language in all its complexity (Newport 1990). Adults might not be able to do this, as they have already constrained their language learning space by much data and experience. Thus, statistical, or incremental learning might start as an application of domain-general learning algorithms, but once training within a domain has progressed, a human or neural network would henceforth be constrained by their domain-specific learning function.

Unfortunately, Elman's original findings have seen limited success in replication in the neural network domain (Rohde & Plaut 1999, 2003). There is, however, recent evidence from AL experiments with human participants that has found that externally constrained incremental learning better facilitates learning of complex embedded AL structures (Poletiek, Conway, Ellefson, Lai, Bocanegra & Christiansen 2018), although there is always a general trend of lower performance at progressively higher levels of embedding. These researchers do not follow Elman's dismissal of externally imposed incremental learning, arguing that a lot of infant-directed speech is simpler in form.

Other studies have found the same effect, and have also discovered that initial exposure to adjacent or non-embedded exemplars is necessary for incremental learning to apply (Lai & Poletiek 2011). In a similar vein, semiartificial learning experiments have demonstrated implicit learning of recursive structures, and that they can be subsequently adapted to higher levels of recursion (Tao & Williams 2018). This could still be an example of discovering or activating a rule for recursion, which is more in line with generative assumptions, although it could also be an example of a domain-general process of incremental learning.

RNNs have also been shown to be able to handle NAD structures in natural language, such as subject-verb agreement in English, but this required explicit grammatical target training (Linzen, Dupoux & Goldberg 2016). Another experiment, however, has found that an RNN that is only trained as a general language model can almost match human performance for subject-verb agreement in Italian, even when nonce words are used, thus suggesting that acquisition of deeper syntactic structure has occurred (Gulordava, Bojanowski, Grave, Linzen & Baroni 2018). Further and more conclusive neural network experiments with data from natural language could strengthen the domain-general argument, as some of the methodological issues with AL experiments are bypassed, though it is still possible that neural network learning algorithms and human learning are not exactly equivalent processes.

5 Language and the Brain

Another salient issue for the discussion of domain-generality and domain-specificity is the research on syntactic processing in the brain. Are there areas or activation patterns of the brain that are unique to language, even for syntax, or do they overlap with other domains?

Traditional neurolinguistic models have often followed the Wernicke-Lichtheim-Geswind (WLG) model, which situates the language faculty responsible for syntactic processing in the left perisylvian cortex, with comprehension centred in the temporal lobe (Wernicke's area), and production in the frontal lobe (Broca's area) (Hagoort 2013). The equivalent system has been observed in non-human brains as well, for example, in songbirds' song processing and production faculty (Moorman, Gobes, Kuijpers, Kerkhofs, Zandbergen & Bolhuis 2012).

Interestingly, Broca's area has also been found to have modular subdivisions within the linguistic domain. An experiment (Goucha & Friederici 2015) which isolated input to contain no semantic cues, nor any derivational morphology, has found that while there is broad activation within the left hemisphere in full, semantically complete sentences (Brodmann areas 44/45/47), only Brodmann area 44 is active in these purely syntactic input conditions (pp. 299-300). This has been taken as evidence for the modularity of language processing in the brain, with syntax being processed separately from semantics, which further strengthens a hypothesis of neurological innate capabilities for syntax.

However, the traditional WLG model has been questioned. Hagoort (2013) conceptualises the function of Broca's area to be less restrictive than domain-specific linguistic processing, instead defining it as a general assembler of linguistic structure. This includes structures that violate phrase structure grammar, as experiments with ALs limited to sequentially ordered recursion show activation for this area in the same way as natural syntax (Petersson, Folia & Hagoort 2012). Even by simply following the WLG model, there is evidence for Broca's area being involved in more cognitive domains than syntactic processing. Broca's area has been found to be activated for music processing, with damage in this area resulting in both problems with syntax comprehension, and music perception (Patel 2003). There are indeed neurological and structural correlates between language and music processing in multiple areas (Jäncke 2012), which the attempts to formulate a formal generative grammar of music in analogy to the hierarchical structure of generative language (Katz & Pesetsky 2009, Lerdahl & Jackendoff 1983) might have illustrated.

Alternatively, as both language and music involve complex relationships between sequences, it has been suggested that Broca's area is involved with domain-general sequence learning, and that no domain-specificity or modularity, such as a language faculty applying Merge, needs to be defined (Christiansen & Chater 2015: 4-5). The human brain seems to process hierarchically complex sequential input including language in fairly similar ways (see, e.g., Forkstam, Hagoort, Fernandez, Ingvar & Petersson 2006, as these processes all activate Broca's area amongst others, but there does seem to be some modality-specificity related to sensory input; particularly phonological data might be processed differently from other kinds, Conway & Pisoni 2008).

The human ability for complex sequence learning might also have a genetic correlate. The FOXP2 gene has been shown to be related to sequence learning abilities in humans, as mutation of this gene can result in both speech and orofacial motor impairments. Two amino acid changes occurred after humans and chimpanzees split evolutionarily (Christiansen & Chater 2015: 4); this could then possibly be an explanation for differences in language learning abilities between humans and other primates (see, e.g., Newport et al. 2004).

A recent experiment using functional near-infrared spectroscopy (fNIRS) data has found, however, that 2-year-old children can detect linguistic NAD violations, but not 3-year-olds, while the reverse is true for purely tonal NAD data, thus suggesting differences in neurological processing between these domains (van der Kant, Männel, Paul, Friederici, Höhle & Wartenburger 2020). There is also counterevidence for music being a separate system from language, coming from a case of amusia where the patient retained speech and rhythm processing abilities (Piccirilli, Sciarma & Luzzi 2000).

Another potential problem with many of the studies that suggest equivalent neural sequence learning and linguistic processes, is that they in some cases use sequence learning stimuli which either diverge structurally very little from generative syntax (Forkstam et al. 2006, Petersson et al. 2012), or involve a phonologically permissible grammar based on simpler formal systems than what Chomsky defines for human language (Christiansen, Conway & Onnis 2012, Christiansen, Kelly, Shillcock & Greenfield 2010). While these experiments do provide evidence for the argument that the areas of the brain involved in syntactic processing should not be restricted to simply being the language faculty as generative theory currently describes it, a looser theory for the language faculty that is constrained to any simple recursive operation, and which also includes simpler grammatical operations in the speech modality, might still be possible. Nevertheless, it is unclear what would remain of any function of an innate language faculty if it was further weakened from its incarnation as Merge.

6 Discussion and Conclusion

In this paper, I have reviewed literature on whether fundamental principles of generative syntax such as NADs and recursion can be accounted for solely by domain-general processes. I started by laying out these principles as they arose in syntactic theory. Chomsky himself has in later years narrowed the strictly universal element of UG to recursion, yet recursion is still not necessarily universal, neither within individual languages, nor crosslinguistically. Nevertheless, even if it is restricted, both recursion and NADs occur prevalently in human language. Is the ability to process these structures then domain-general or domain-specific?

Statistical learning experiments, using ALs of a finite state grammar design, have effectively demonstrated quick learning of adjacent dependencies which might inform syntactic knowledge, but the evidence for NAD acquisition in this paradigm is subject to further constraints, and this finite state grammar AL design does not allow for recursion, which reduces its veracity for describing natural syntactic diversity. The incremental learning approach arising from Elman's (1993) RNN research shows greater success, even with complex centre-embedded structures, also when it has been extended to the AL domain with human participants. Nevertheless, neural networks are not humans, and Elman himself criticised the validity of external constraints on incremental learning as a model for child language acquisition. It is also possible that domain-general statistical learning is combined with UG in the brain, given the success of some probabilistic phrase structure grammars.

I also discussed the neurological evidence for domain-generality. While there is evidence for modular and language-specific activation of the brain, including for syntax in Broca's area, there is also evidence for the same areas being activated for other sequence processing and learning tasks, which together with some genetic evidence could suggest that while human brains are adapted for complex sequential processing, this is not a strictly domain-specific ability. However, the sequence learning evidenced in these experiments could still fit within a looser language faculty interpretation.

Indeed, a fundamental problem with trying to rule out any domain-specific component to syntactic processing is that the definition of the language faculty could continue to get broader in scope, broader in which modalities it allows, as evidenced by the idea of generative music. A question to keep in mind here might be whether either approach provides a simpler explanation of observed phenomena than the other. If domain-general sequence learning abilities are shown to perfectly converge with natural language syntax learning, both in learning outcomes and in brain activation, then it would arguably be a simpler explanation of overall human learning abilities.

In further research, this could be evidenced by comparing learning and brain activation for recursive and NADs in both naturalistic ALs and non-linguistic equivalents. The same ALs should also be given to a neural network, following a similar, possibly incremental, training strategy, in order to assess the abilities of a purely statistical learner.

References

Abdi, H. 1994. A neural network primer. *Journal of Biological Systems* 02(03). 247–281.

Chomsky, N. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2(3). 113–124.

Chomsky, N. 1959a. A review of BF Skinner's Verbal Behavior. *Language* 35(1). 26–58.

Chomsky, N. 1959b. On certain formal properties of grammars. *Information and Control* 2(2). 137–167.

Chomsky, N. 1981. *Lectures on Government and Binding: The Pisa lectures.* Walter de Gruyter.

Chomsky, N. 1995. *The minimalist program.* MIT Press.

Christiansen, M. H. & N. Chater. 2015. The language faculty that wasn't: a usage-based account of natural language recursion. *Frontiers in Psychology* 6(1182).

Christiansen, M. H., C. M. Conway & L. Onnis. 2012. Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes* 27(2). 231–256.

Christiansen, M. H., M. L. Kelly, R. C. Shillcock & K. Greenfield. 2010. Impaired artificial grammar learning in agrammatism. *Cognition* 116(3). 382–393.

Conway, C. M. & D. B. Pisoni. 2008. Neurocognitive Basis of Implicit Learning of Sequential Structure and Its Relation to Language Processing. *Annals of the New York Academy of Sciences* 1145(1). 113–131.

Culbertson, J., P. Smolensky & C. Wilson. 2013. Cognitive Biases, Linguistic Universals, and Constraint-Based Grammar Learning. *Topics in Cognitive Science* 5(3). 392–424.

Elman, J. L. 1991. Distributed Representations, Simple Recurrent Networks, And Grammatical Structure. *Machine Learning* 7(2/3). 195–225.

Elman, J. L. 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48(1). 71–99.

Evans, N. & S. C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–448.

Everett, D. L. 2005. Cultural Constraints on Grammar and Cognition in Pirahã. *Current Anthropology* 46(4). 621–646.

Fitch, W. T. & M. D. Hauser. 2004. Computational Constraints on Syntactic Processing in a Nonhuman Primate. *Science* 303(5656). 377–380.

Forkstam, C., P. Hagoort, G. Fernandez, M. Ingvar & K. M. Petersson. 2006. Neural correlates of artificial syntactic structure classification. *NeuroImage* 32(2). 956–967.

Gebhart, A. L., E. L. Newport & R. N. Aslin. 2009. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review* 16(3). 486–490.

Gómez, R. L. 2002. Variability and Detection of Invariant Structure. *Psychological Science* 13(5). 431–436.

Goucha, T. & A. D. Friederici. 2015. The language skeleton after dissecting meaning: A functional segregation within Broca's area. *NeuroImage* 114. 294–302.

Gulordava, K., P. Bojanowski, E. Grave, T. Linzen & M. Baroni. 2018. Colorless Green Recurrent Networks Dream Hierarchically. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, Association for Computational Linguistics.

Hagoort, P. 2013. MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology* 4.

Hauser, M. D., N. Chomsky & W. T. Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298(5598). 1569–1579.

Hauser, M. D., E. L. Newport & R. N. Aslin. 2001. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78(3). B53–B64.

Jäncke, L. 2012. The Relationship between Music and Language. *Frontiers in Psychology* 3(123).

van der Kant, A., C. Männel, M. Paul, A. D. Friederici, B. Höhle & I. Wartenburger. 2020. Linguistic and non-linguistic non-adjacent dependency learning in early development. *Developmental Cognitive Neuroscience* 45. 100819.

Katz, J. & D. Pesetsky. 2009. The identity thesis for language and music http://ling.auf.net/lingBuzz/000959.

Kirkham, N. Z., J. A. Slemmer & S. P. Johnson. 2002. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83(2). B35–B42.

Lai, J. & F. H. Poletiek. 2011. The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition* 118(2).

265–273.

Lerdahl, F. & R. S. Jackendoff. 1983. *A Generative Theory of Tonal Music.* MIT Press Ltd.

Linzen, T., E. Dupoux & Y. Goldberg. 2016. Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *Transactions of the Association for Computational Linguistics* 4. 521–535.

Moorman, S., S. M. H. Gobes, M. Kuijpers, A. Kerkhofs, M. A. Zandbergen & J. J. Bolhuis. 2012. Human-like brain hemispheric dominance in birdsong learning. *Proceedings of the National Academy of Sciences* 109(31). 12782–12787.

Newport, E. L. 1990. Maturational Constraints on Language Learning. *Cognitive Science* 14(1). 11–28.

Newport, E. L. & R. N. Aslin. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48(2). 127–162.

Newport, E. L., M. D. Hauser, G. Spaepen & R. N. Aslin. 2004. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* 49(2). 85–117.

Patel, A. D. 2003. Language, music, syntax and the brain. *Nature Neuroscience* 6(7). 674–681.

Perruchet, P. & A. Rey. 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review* 12(2). 307–313.

Petersson, K.-M., V. Folia & P. Hagoort. 2012. What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language* 120(2). 83–95.

Piccirilli, M., T. Sciarma & S. Luzzi. 2000. Modularity of music: evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery & Psychiatry* 69(4). 541–545.

Pinker, S. & P. Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13(4). 707–727.

Poletiek, F. H., C. M. Conway, M. R. Ellefson, J. Lai, B. R. Bocanegra & M. H. Christiansen. 2018. Under What Conditions Can Recursion Be Learned? Effects of Starting Small in Artificial Grammar Learning of Center-Embedded Structure. *Cognitive Science* 42(8). 2855–2889.

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

Reinhart, T. M. 1976. *The syntactic domain of anaphora*: Massachusetts Institute of Technology dissertation.

Robinson, P. 2005. Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition* 27(02).

Rohde, D. L. & D. C. Plaut. 1999. Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition* 72(1). 67–109.

Rohde, D. L. & D. C. Plaut. 2003. Less is Less in Language Acquisition. In *Connectionist models of development*, 178–218. Psychology Press.

Saffran, J. R. 2001. Words in a sea of sounds: the output of infant statistical learning. *Cognition* 81(2). 149–169.

Saffran, J. R. 2002. Constraints on Statistical Language Learning. *Journal of Memory and Language* 47(1). 172–196.

Saffran, J. R., R. N. Aslin & E. L. Newport. 1996. Statistical Learning by 8-Month-Old Infants. *Science* 274(5294). 1926–1928.

Saffran, J. R., E. K. Johnson, R. N. Aslin & E. L. Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1). 27–52.

Tallerman, M., F. Newmeyer, D. Bickerton, D. Bouchard, E. Kaan & L. Rizzi. 2009. What Kinds of Syntactic Phenomena Must Biologists, Neurobiologists, and Computer Scientists Try to Explain and Replicate? In D. Bickerton & E. Szathmáry (eds.), *Biological Foundations and Origin of Syntax*, 135–157. The MIT Press.

Tao, Y. & J. N. Williams. 2018. Generalization of Syntactic Knowledge in Semiartificial Language Learning. *Language Learning* 68(4). 1001–1031.

Yang, C. D. 2004. Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10). 451–456.

Zettersten, M., C. E. Potter & J. R. Saffran. 2020. Tuning in to non-adjacencies: Exposure to learnable patterns supports discovering otherwise difficult structures. *Cognition* 202. 104283.

Dag-Tore Nordbø Kristiansen
University of Cambridge
dtnk2@cam.ac.uk