Artificial language learning

Jennifer Culbertson

February 14, 2021

Abstract

Artificial language learning experiments, first used as early as the 1950s, have helped language acquisition researchers answer longstanding questions about how learners derive representations and make generalizations based on exposure to limited data. Recently, they have been co-opted by theoretical linguists to test hypotheses about how properties of human cognition shape natural language phonology, morphology, and syntax. Empirical evidence derived from these methods has been used to build more precise accounts of the link between how languages are learned (and processed) and cross-linguistic tendencies long-noted in the typological record. This chapter explains why artificial language learning is an important tool in the syntactician's toolbox, what phenomena it has been used to study to date, and where research with these methods is heading in the future.

1 Introduction

The earliest uses of artificial language (or grammar) learning in psychology were focused on whether learners could extract implicit rules or rule-like generalizations from structured input (e.g., Braine, 1963; Reber, 1967). These experiments were extended starting in the 1990s to show that learners could use distributional information in the input—statistical learning—to form representations of word boundaries and phrasal constituents, along with syntax-like rules (Saffran et al., 1996; Mintz et al., 2002; Reeder et al., 2013). More recently, researchers in cognitive psychology and theoretical linguistics have begun adapting artificial language learning methods to study how statistical learning might interact with or be shaped by other properties of the cognitive system. An explicit goal of this research, and the focus of this chapter, is using these methods to provide evidence for cognitive constraints or biases which might explain specific features of language structure.

Such constraints form a cornerstone of generative theories of grammar, where they are posited in order to restrict the set of possible human languages. The two main traditional sources of evidence for grammatical constraints are developmental pathways during acquisition, and common features of synchronic and diachronic typology. For example, common errors in the acquisition of phonology, such as the simplification of multisyllabic words or consonant clusters, have been used as evidence for positing universal constraints on syllable structure (e.g., Barlow 2000). These same constraints are supported to the extent that they can generate all and only typologically attested syllable structure patterns (Prince & Smolensky, 1993/2004). Similarly, research in syntax has looked for evidence of universal principles, like structure-dependence, both in the types of errors children make (or fail to make) during early acquisition and in the types of rules found (or not found) in the world's languages (e.g., Crain & Nakayama, 1987).

1.1 Empirical and theoretical challenges to universal constraints

The view of language, and language learning, as subject to a set of universal constraints is challenged both by empirical facts and alternative theoretical approaches. Here I will argue that artificial language learning experiments can be used both to test predicted effects of hypothesized constraints, and to refine how the constraints should be formalized in a theory of grammar. First, however, I will briefly lay out some of the challenges linguists face in making a convincing case for particular constraints on language using natural language acquisition and typological data alone.

Perhaps the most well-known argument for constraints on the linguistic system comes from the 'argument from the Poverty-of-the-Stimulus' (Chomsky, 1959). The general sketch of this argument is that what children acquire—abstract, generalizable grammatical knowledge—and how they acquire it—with remarkable speed, without explicit guidance, and with markedly better results than adults—is only possible with a set of universal guiding principles in place. Of course, almost all aspects of the argument have been vigorously challenged (e.g., Morgan et al., 1995; Pullum & Scholz, 2002; Ambridge et al., 2008), and at the very least we now have a much clearer picture of how much grammatical knowledge learners can in fact acquire on the basis of the input they get. Computational models of phonotactic learning, word segmentation, and syntactic category learning, for example, have been built to explore this in mathematically rigorous ways (e.g., Hayes & Wilson, 2008; Frank et al., 2010; Redington et al., 1998, among many others). That does not necessarily mean that learning is not constrained (Legate & Yang, 2002), however it does mean that evidence of sophisticated, early-acquired knowledge on its own may not indicate the presence of universal guiding principles. This has important implications for research aimed at showing that certain logically possible structures or rules are not learned, that there are differences in the rate of learning among alternative patterns, or that certain error types are more common than others. In these cases, one must show quantitatively that these asymmetries are not present in, or otherwise driven by particular properties of the input.

Research highlighting the power of statistical or distributional learning mechanisms may suggest that a grammar can be acquired on the basis of the input alone, largely obviating the need to posit constraints on learning. A potentially strong counter to this is the existence of typological universals. If certain types of linguistic patterns, rules, or structures are systematically missing from the typological record, then perhaps we don't want our learning mechanism to be able to acquire them, even if it could. In other words, we may need to posit a constraint or principle which actively rules them out. On the face of it, typological universals therefore provide compelling evidence of constraints on language. These constraints could be active during learning in particular, or could persist in later language usage, shaping how linguistic systems change over time. There are two main challenges linguists face in inferring constraints on language directly from the typological record: limitations on the data itself, and challenges involved in interpreting reliable typological differences. Typological universals are formulated on the basis of a sample of languages, in some case very small (for example, Greenberg's sample of 30), but more recently relatively large (e.g., 2,679 in Dryer & Haspelmath, 2013)). Although larger samples may result in a more reliable picture of what types of linguistic systems are attested, any sample will necessarily omit many hypothetically possible languages which simply don't exist. This makes it difficult if not impossible to establish to a satisfactory level of certainty that a universal is absolute, or inviolable (Piantadosi & Gibson, 2014). Further, samples are confounded by genetic and areal relations among the languages–in other words individual languages cannot be treated as independent data points (Cysouw, 2005). There are computational methods which make it possible to take these factors into account (e.g., Bickel, 2007; Dunn et al., 2011), however they are not without limitations (Jaeger et al., 2011).

More importantly, even if a typological universal does appear to be very reliable, this does not tell us that the best explanation for it is a principle of grammar, or even a more general property of our cognitive system. This exactly parallels acquisition (where evidence for the absence of particular types of linguistic errors in development does not necessarily implicate a principle of grammar). There is a large body of research in typology as well as theoretical linguistics which argues that most if not all universals result from pathways of diachronic change, not constraints active in the cognitive systems of individuals. For example, evolutionary phonology argues that many phonology universals results from 'channel bias'—misperceptions or misarticulations which over time change languages in systematic ways. In other words, these patterns reflect constraints on the physical/sensory system rather than from cognitive or grammatical principles shaping learning (Ohala, 1992; Blevins, 2004). Similar work argues that grammaticalization pathways in syntax lead to universals relating to morphosyntax and word order (among other things, Aristar, 1991; Bybee, 2006; Whitman, 2008).

This is all in addition to the fact that many proposed universals are in fact known to admit exceptions, i.e., they are statistical tendencies rather than absolute universals. This has led some researchers to reject the notion of domain-specific constraints on language (e.g., Evans & Levinson, 2009), arguing instead that general cognitive constraints largely determine likely grammatical systems. Indeed, statistical universals are not easily integrated into many generative frameworks, which formalize principles of grammar as inviolable constraints (or in approaches like Optimality Theory, a universal set of violable constraints which strictly limits the generative capacity). This leads to a serious theoretical dilemma: should theories of natural language structure account only for absolute universals, or should they incorporate a formal notion of preference or bias in order to account for these tendencies? The former approach is assumed in most theories of generative syntax. For example Newmeyer (2005) argues explicitly that statistical tendencies be accounted for by grammar-external factors. These might include both historical and culturally-driven pressures, as well as 'third factor' constraints on cognition (Chomsky, 2005). However, several recent models working within a probabilistic constraint-based grammatical framework take the latter approach to at least some statistical typological tendencies (Pater, 2011; Culbertson et al., 2013; White, 2017).

To summarize then, although typological universals present a potentially strong piece of evidence for constraints on the linguistic system, the data we get from language samples is not necessarily reliable, and may in some cases be misleading. Nor can we say with confidence that a universal which does appear to be reliable should be explained by features of our linguistic or cognitive system. For some researchers, this problem is exacerbated by the fact that many typological tendencies to not conform to the notion of exceptionless linguistic universal argued for by many generative syntacticians.

1.2 How can artificial language learning methods help?

The most important contribution of artificial language learning experiments to date is in allowing researchers to test the predicted behavioral effects of hypothesized constraints in a controlled laboratory environment. Observations from linguistic typology, or language acquisition can be used to generate hypotheses linking language structure to human cognition. The predictions of these hypotheses can be tested using precisely designed experimental manipulations. While most work in theoretical syntax does not yet incorporate this kind of evidence, the last decade has seen a surge in the use of artificial language learning experiments in research on theoretical phonology. For the most part, these studies have explicitly focused on statistical typological tendencies, attempting to show that typologically common patterns are acquired (or inferred) more readily than rare patterns. In other words, this research is focused on uncovering cognitive biases which might explain a given typological distribution. Such biases are difficult to test with natural language acquisition data alone, since no two natural languages will differ only in the phenomenon of interest. Further, the researcher cannot control the frequency with which particular learners might receive relevant information in the input. Where research has in fact focused on apparently non-defeasible principles, the problem is obvious: there simply are no natural language acquisition data available.

Using artificial language learning experiments makes it possible to perfectly match languages aside from properties of interest, to control input frequency, and to compare learning of attested versus unattested and common versus rare linguistic patterns. In addition to this, these methods allow us to explore whether the same biases are found across development, how they might be amplified or dampened by language experience, and how widely they apply across cognitive domains. In the remainder of this chapter, I will first outline four general artificial language learning methods used for research on syntax. I will then give some specific examples of results obtained using these methods, briefly making connections to similar work in phonology where relevant. Because artificial language learning is relatively new in the field of syntax, I will end by discussing the directions this research must head in the future to fulfill its potential to answer key questions for syntactic theory.

2 Key methods and results

2.1 A description of four widely used artificial language learning paradigms

Table 1 provides a brief summary of the four main paradigms I will discuss in this chapter, along with a selection of key references. These paradigms differ in terms of what the input looks like and what the behavioral measure of interest is. It should be noted that although I will continue to refer to them collectively as artificial language learning paradigms, they do in fact differ in the extent to which 'learning' is actively involved. The first paradigm, which I call 'ease of learning' is perhaps the most basic one. Participants are typically taught one of a set of patterns of interest, and are tested on how accurately (within some set number of trials) they are able to learn it. Accuracy levels are then compared across patterns. For example, Tabullo et al. (2012) teach participants a language with one of the six basic word order patterns, and test which of the six patterns is learned better. A variant of this paradigm traces the learning trajectory over trials, in some cases without an explicit training block, in order to assess how quickly a given pattern is acquired.

The 'poverty-of-the-stimulus paradigm' was first used in Wilson (2003, 2006), to explore biases in learning of phonological alternations. As its name indicates, the paradigm exposes participants to impoverished input, missing evidence that would support one hypothesis over another. For example, the input may be ambiguous between two grammatical rules, with held-out data designed to allow researchers to observe whether participants inferred one rule or the other. To take an example from syntax (discussed in detail below), Culbertson & Adger (2014) train participants on nouns with single modifiers (e.g., adjectives or numerals), leaving ambiguous the relative order of modifiers when more than one is present. Thus it differs from a basic ease of learning design in using held-out data to test how learners trained on a subset of the artificial language extrapolate to structures of interest in the absence of explicit evidence. Note that this is not simply testing generalization to new stimuli of the same kind (something which is commonly done across paradigms), but to patterns or structures which differ from exposure stimuli in some critical way. In more recent work in syntax, this paradigm has instead been called the 'extrapolation paradigm' (in nod to this distinction, and incidentally also disassociating the paradigm from the contentious POS debate, e.g., Maldonado & Culbertson, in press).

The 'regularization' paradigm involves training learners on a variable or probabilistic distribution of alternative structures in the input–for example two possible words for a given meaning, or two different word orders used to describe a given scene. The input language is thus similar to what a learner might encounter during a period of linguistic innovation, contact, or change. Evidence for a bias comes from observing if and how learners alter the input distribution in their output when tested. Using one of the alternatives more frequently than would be justified by the input is known as 'regularization'. Generally speaking, regularization has been used as an alternative measure of ease of learning, or preference; patterns that are regularized are taken to be preferred by learners. For example, Culbertson et al. (2012) teach participants variable systems of noun phrase order (adjectives and numerals come before or after the noun with some probabilitity), and test whether learners shift these variable systems to bring them in line with a proposed preference for word order 'harmony' (i.e. consistent order of heads and different modifiers).

The fourth paradigm, called the 'silent gesture' paradigm, involves eliciting manual gestures, typically in the context of improvisation (with no input) rather than learning (thus it is essentially an extreme version of the 'poverty-of-the-stimulus' paradigm, although see Motamedi et al., 2019). Hearing participants with no previous knowledge of a sign language are given a set of meanings which they must communicate using gesture alone, without any speech. Researchers measure similarities among the gestures they produce, as an indication of constraints or biases which guide behavior when no input or linguistic model is present. This paradigm was first used by Goldin-Meadow et al. (2008) who observed similarities in gestures conveying simple transitive events among speakers of several different languages.

Table 1: Summary of key artificial language learning methods.

Paradigm	Brief description	Key references
Ease of learning	Learners are trained on patterns of interest. Speed or accuracy of learning is compared across patterns.	Morgan et al. (1989), Musso et al. (2003), Culbertson et al. (2017), Maldonado & Culbertson (in press)
Poverty-of-the-stimulus (Extrapolation)	Learners are trained on input data that is ambiguous between hypotheses of interest. Extrapolation to disambiguating hold-out data is observed.	Wilson (2006), Culbertson & Adger (2014), Martin et al. (2020), Maldonado & Culbertson (in press)
Regularization	Learners are trained on variable data. Degree of regularization of input distributions is observed (or compared across patterns).	Hudson Kam & Newport (2009), Culbertson et al. (2012), Fedzechkina et al. (2012), Culbertson et al. (2020a)
Silent gesture	Typically, no input is provided. Participants improvised gestures or gesture sequences are observed.	Goldin-Meadow et al. (2008), Schouwstra & de Swart (2014), Futrell et al. (2015a), Culbertson et al. (2020b)

2.2 Some methodological notes

Before heading into discussion of specific experiments and findings using artificial language learning, it is worth highlighting a couple of general methodological issues. First, because this is a relatively new area of inquiry, the methods are still under development to some degree. For example, the size and complexity of the artificial languages used varies widely—from studies using many novel lexical items, and exposure over several days (Hudson Kam & Newport, 2009; Fedzechkina et al., 2012), to studies using native or pseudo-native lexical items and lasting 10-20 minutes (Smith & Wonnacott, 2010; Culbertson & Adger, 2014). Whether these methodological differences matter is not well understood. In addition, as I will discuss further below, the populations targeted by these experiments are often limited to adults, who necessarily bring their experience with their native language to the task. The specific language has historically often been English (an issue shared with (psycho)linguistics, psychology, and indeed cognitive science more generally), although more recent work includes evidence from other populations. Setting aside experience with a particular language, that this research mainly targets adult participants is also important to note in the context of broader questions about the hypothesized linking mechanisms between individual-level cognitive biases that play out in the lab, and population-level forces which lead to typological distributions. Specifically, there is robust debate in the field as to whether the biases of adults or children, L1 learners/speakers or L2 learners/speakers, primarily shape language typology (e.g. see, Lightfoot, 1997; Bybee, 2009; Lupyan & Dale, 2010, among others). This issue is beyond the scope of this chapter, but should be kept in mind by the interested reader. Finally, it is worth mentioning that, like all experimental methodologies, artificial language learning is by design very much simplified compared to the real-world equivalent. This has, in my experience, led to much skepticism that these methods could in fact tap into mechanisms that are at play in natural language (learning). As I have outlined above, however, there are important limitations to all sources of data in linguistics, and therefore converging evidence is extremely valuable to the field. To the degree that these sources of evidence align, we can be confident that the results we see are meaningful.

2.3 Early and foundational studies

The first artificial language (or 'grammar') learning experiments, such as those conducted by Braine (1963) and Reber (1967), were aimed at understanding whether language was acquired via implicit learning, as argued by Chomsky and other early generativists (Chomsky, 1957). The 'languages' in most early artificial language studies were comprised of strings of nonsense words generated by a finite-state grammar, with no corresponding semantic content.[1] Later studies began using semantically meaningful strings, following Moeser & Bregman (1972, 1973) who showed that syntactic rules are much more readily acquired under these conditions. The general idea was to observe how learners gradually pick up on the underlying structure of the language over the course of the experiment. Of particular interest was whether participants who accurately learned the system could verbalize their knowledge, or whether it remained implicit. Indeed, these early studies have prompted a huge amount of further work on implicit learning of grammatical rules (e.g., Mathews & Roussel, 1997; Williams, 2005). Braine and colleagues also went on to use these methods to explore implicit learning of grammatical categories including morphological gender. For example, Braine et al. (1990) and Brooks et al. (1993) showed that adults and children can acquire artificial gender classes when phonological cues are present, but not when classes are completely arbitrary. This was a theoretically significant finding, since early models of morphological acquisition predicted easy learning of arbitrary subclasses (e.g., Anderson, 1983; Pinker, 1984).

While most of these early studies focused on general features of the learning mechanisms involved in acquiring grammatical knowledge, researchers were already beginning to compare learning of different types of systems. Frigo & McDonald (1998) compared ease of learning of artificial gender systems depending on the position and weight of phonological cues to noun class. A series of studies by Newport and colleagues (Morgan & Newport, 1981; Morgan et al., 1987, 1989) showed

[1] It's worth noting in relation to this that there is huge literature on the use of artificial language learning paradigms to explore the ability of humans and other animals to acquire *center-embedding*—sometimes argued to be a key feature of language, and something only humans can acquire. This claim is very contentious, and it turns out center-embedding is very hard to learn, at least in the context of meaningless strings of nonsense words. For reasons of space, I will not discuss these studies further, but point the interested reader to: Fitch & Hauser (2004); Perruchet & Rey (2005); de Vries et al. (2008); Lai & Poletiek (2011), among many others.

that learning artificial phrase structure grammars was facilitated by cues to constituent structure (e.g., morphological dependencies and prosody). Smith et al. (1993), and later Musso et al. (2003) compared learning of rules from an existing language (e.g., passive formation in Italian), and hypothetical 'impossible' rules (e.g., negation by changing word order).

In a now classic set of studies, Hudson Kam & Newport (2005, 2009) sought to directly observe how learners might change a language in failing to accurately learn it. The idea was that these changes might reveal learners' preferences or expectations about how natural languages should be structured. Research investigating how deaf children acquire sign language from second-language learner caregivers (Singleton & Newport, 2004) suggested that unstructured errors in the input were not reproduced by children. Where adults sometimes produced a correct form, but other times omitted required morphological markers, children appeared to acquire deterministic rules. Regularization of unstructured variation has also been documented in research on creolization, a common claim being that child learners reduced variability during the process of creolization (DeGraff, 1999; Singler, 2006). Hudson Kam & Newport (2005, 2009) developed the 'regularization' paradigm, to investigate this phenomenon in the lab. Adult and child participants were trained on languages in which a determiner was used unpredictably (e.g., one language had the determiner present in most cases, but absent in the remainder). The frequency with which learners used the determiner in subsequent productions was measured. The findings revealed that children tended to produce systems which were more consistent, or regular than the input, while adults only did this when the input was particularly complex. This regularization behavior (sometimes also measured as a drop in entropy compared to the input) has now been investigated extensively to show how children and adults might push linguistic and non-linguistic systems to be more deterministic over generations of learners (Hudson Kam & Chang, 2009; Reali & Griffiths, 2009; Smith & Wonnacott, 2010; Culbertson et al., 2012; Perfors & Burns, 2010; Ferdinand et al., 2019; Saldana et al., in press).

While the regularization paradigm was originally designed to investigate learners' expectations about linguistic variation, experimental paradigms developed at the same time by theoretical phonologists focused on investigating the factors influencing the content and structure of phonological rules. For example, influential studies by Wilson (2003, 2006), Pycha et al. (2003), and Moreton (2008) used the poverty-of-the-stimulus and ease of learning paradigms to test competing hypotheses from two alternative theoretical approaches. Evolutionary phonology argues that typological generalizations reflect pathways of phonologization, common articulatory and/or perceptual errors which accumulate over generations (Ohala, 1993; Blevins, 2004). By contrast, substantively biased approaches to phonological theory argue that typology reflects biases in the grammatical systems of individuals (Steriade, 1997; Hayes, 1999). The latter predicts that the effects of individual biases should be apparent in the learning of phonological patterns, straightforwardly testable using artificial language learning. These two strands of research–from developmental psychology on the one hand, and theoretical phonology on the other–built the foundation for experimental work investigating specific biases hypothesized to shape properties of syntactic systems.

In the next sections, I will focus on three main categories of phenomena: simplicity (syntactic patterns reflecting a preference for representational complexity), naturalness (patterns reflecting meaning or conceptual knowledge), and communicative efficiency (patterns reflecting a trade-off between simplicity and ambiguity avoidance). I will end with discussion of some additional studies on processing and perception. As noted above, the majority of these studies target adult learners, but see Culbertson & Schuler (2019) for a review focused on relevant artificial language learning research in children.

2.4 Simplicity

Observationally, linguistic systems which are formally less complex are often more common than more complex alternatives. This has been incorporated into linguistics theories, particularly in phonology (Martinet, 1968; Clements, 2003; Prince & Smolensky, 1993/2004), and has been proposed as a general inductive bias in the broader cognitive science literature (e.g., Chater & Vitányi, 2003). In generative syntax, economy principles have been claimed to constrain representations and operations (e.g., Chomsky, 1957; Chomsky & Lasnik, 1993; Grimshaw, 1997). Simplicity is also a key notion in definitions of communicative efficiency (see Section 2.6 below). A number of artificial language learning studies have sought to provide behavioral evidence that simpler structures are easier to learn, or more likely to be implicitly assumed by learners in the absence of evidence in the input. For example, ease of learning of artificial phonological patterns generally correlates with the number of features relevant to the pattern (see Moreton & Pater, 2012, for a review of such studies in phonology). In syntax, artificial language learning has been used to investigate one of the most well-known observations concerning word order, alternatively called harmony or consistent head-direction (Greenberg, 1963; Travis, 1984; Dryer, 1992; Baker, 2001; Hawkins, 2004; Cinque, to appear), which a number of researchers have argued to reflect simplicity.

2.4.1 Word order harmony

Harmonic word order patterns are those which preserve a consistent order of syntactic heads relative to modifiers and other dependents, across a range of phrase types. For example, a language which has Object-Verb order in the VP, Noun-Postposition order in the PP, and Adjective-Noun order in the NP, is harmonic (head-final) across these phrases. A number of alternative mechanisms have been proposed to explain harmony, including representationally simplicity (e.g., Vennemann, 1976; Pater, 2011; Culbertson & Newport, 2015; Culbertson & Kirby, 2016), processing ease (e.g., Hawkins, 1994), a head-direction 'parameter' (e.g., Travis, 1984), common grammaticalization pathways (e.g., Whitman, 2008) or accidental, lineage-specific effects (e.g., Dunn et al., 2011). As a general tendency with many exceptions, harmony is thus subject to the issues of interpretation highlighted above (see also Ladd et al., 2014): it is likely to reflect a bias rather than a hard-and-fast constraint, but based on the typology alone we cannot say with certainty that any cognitive (or linguistic) bias for harmony exists.

Culbertson et al. (2012) sought to investigate the cognitive underpinnings of word order patterns in the nominal domain, including harmony among noun phrases with an adjective and numeral modifier. In a relatively large sample of languages (Dryer, 2013a,b), harmonic orders (e.g., both modifier types pre-nominal, or both post-nominal) outnumber non-harmonic orders (e.g., one modifier pre-nominal, and the other post-nominal). In addition, among both harmonic and non-harmonic patterns, post-nominal adjectives are more common: the harmonic pattern N-Adj with N-Num is relatively more common than Adj-N with Num-N, and among the non-harmonic patterns N-Adj with Num-N is more common than Adj-N with N-Num ('Universal 18', Greenberg, 1963). To investigate whether these two tendencies reflect cognitive biases in learning, Culbertson et al. (2012) used the regularization paradigm with adult English speakers. Participants were taught a language with simple noun phrases comprised of a noun and a single modifier, either an adjective or a numeral (i.e., (dis)harmony was at the level of the grammar, not the utterance, cf. Hawkins, 1994). Participants were trained on an input language featuring one of the four patterns (pre- or post-nominal harmonic, N-Adj with Num-N or Adj-N with N-Num) as the dominant one in the language. However, order in each language was variable, and thus in principle any combination of noun with adjective or noun with numeral could appear in either order.

The results showed that participants trained on the two dominant harmonic patterns regularized most, and participants trained on the rare Adj-N, N-Num pattern regularized the least. This result was partially replicated with child learners in Culbertson & Newport (2015), who showed that 6-7 year old children strongly prefer harmonic patterns, and in general shift non-harmonic input distributions dramatically toward harmonic (with no apparent distinction between the two non-harmonic patterns, see also Culbertson & Newport, 2017). Notably, English is a harmonic language, thus a preference for harmony could reflect abstract structural transfer—i.e., the results might reflect prior language experience rather than a universal bias with the potential to shape typology. Culbertson et al. (2020a) therefore used this same design to test French- and Hebrew-speaking children and adults, whose native language pattern is predominantly *non-harmonic* (N-Adj, Num-N). In all cases, a bias for harmony was replicated (though it also interacted in complex ways with both L1 and L2 experience in these populations). Finally, it is worth noting that the notion of harmony is mainly discussed as *cross-category* harmony: alignment across different types of phrases, like VP and PP. Recent research has used the extrapolation paradigm to provide evidence for a harmony bias in this context (Wang et al., submitted).

The experiments outlined above primarily provide evidence of a preference for harmony–a bias likely related to representational simplicity. Follow-up work has sought to understand this as an emergent bias reflecting ease of generalization in grammatical rule learning (Pater, 2011), a prior bias over grammars (Culbertson & Smolensky, 2012; Culbertson et al., 2013), or as a more general bias for simplicity operating across cognition (Culbertson & Kirby, 2016).

While additional evidence is needed both to adjudicate between these theoretical approaches, the original motivation for (Culbertson et al., 2012) was in fact the potential distinction between *non-harmonic* patterns mentioned above (Universal 18). In their study, English-speaking adult learners particularly disfavored combining Adj-N with N-Num (see Culbertson et al., 2013, for a replication of this effect). A possibility explanation for this is that this pattern is the worst of the worst; it is not only non-harmonic, but uses pre-nominal adjectives, which are independently disfavored. For example, post-nominal adjectives may have an advantage over pre-nominal adjective because they set up the object to be modified first, so that the adjective can immediately be interpreted in context (Kamp & Partee, 1995; Rubio-Fernandez et al., 2020). This leads us to another set of experiments exploring the idea that some linear orders may be preferred over others for semantic or conceptual reasons, rather than simplicity.

2.5 Naturalness

The idea that some syntactic structures or order are more 'natural' than others is loosely related to the notion of naturalness commonly appealed to in phonology. In phonology, more natural patterns are phonotactically motivated (e.g., vowel harmony as opposed to disharmony), and a number of artificial language learning studies have sought to explore whether such rules are preferred by learners (e.g., Wilson, 2006; White, 2014; Martin & White, 2019). What constitutes a natural pattern in syntax is not as straightforward, however the terms has been used to refer to patterns which are motivated by semantic or conceptual considerations. There are two strands of research using artificial language learning to explore naturalness biases in syntax: the first investigates basic word order, and the second again tackles word order in the noun phrase.

2.5.1 Basic word order

The typological frequency of basic word order patterns is highly skewed: SOV is the most frequent, followed by SVO, with a steep drop off for VSO, VOS, OVS, and finally OSV. Most theoretical accounts of the typological differences among these orders focus on a general preference to have subjects (or animate things) first (e.g., Gibson, 2000; Jackendoff, 2002; Demiral et al., 2008), and a preference for grouping the verb and the object together (Baker, 2009; Gibson, 2000). The two most frequent orders, SOV and SVO conform to both of these, while the less frequent orders violate one or both. While some psycholinguistic studies suggest these constraints might explain word order variation within a language (e.g., Demiral et al., 2008), typological frequency data alone does not tell us whether these kinds of constraints shape the overall frequency of basic ordering patterns (e.g., see Maurits & Griffiths, 2014).

A handful of artificial language learning studies using an ease-of-learning design have attempted to provide evidence for a cognitive bias favoring frequent basic orders over infrequent ones. Tily et al. (2011) taught English-speaking adults a language featuring one of the six basic word order patterns. Participants were tested on comprehension (choosing which picture corresponded to a given sentence in the language) and production (clicking on words in the lexicon to construct a sentence corresponding to a given picture). No differences in comprehension were found, but production accuracy was higher in the SVO and SOV input conditions compared to the remaining four patterns. Similarly, Tabullo et al. (2012) taught Spanish-speaking adults a miniature artificial language with SVO, SOV, OSV, or VSO word order. They found that accuracy rates (as measured by a grammaticality judgment task) for SVO and SOV were higher than for OSV and VSO, roughly mirroring the typological frequency of these patterns.

Neither of these studies revealed a clear difference in learning between SVO and SOV, despite the frequency difference between them (Greenberg, 1963; Dryer, 1997), and despite the fact that all learners in these studies were speakers of SVO languages. Interestingly, a number of researchers have shown that SOV is more likely than SOV to arise in the early stages of spoken and signed languages, though in some cases SOV later changes to SVO (Givón, 1979; Fischer, 1975; Sandler et al., 2005; Gell-Mann & Ruhlen, 2011). This suggests the possibility that even if there is no learnability difference between these orders, SOV order may be innovated more frequently. An important series of studies explores this possibility using the silent gesture paradigm, first introduced by Goldin-Meadow et al. (2008). In this study, non-signing adult participants who were native speakers of English (SVO), Spanish (SVO), Chinese (SVO/SOV), or Turkish (SOV) were shown pictures of basic transitive events, like a scene in which a girl covers a box. They were asked to communicate these events using only their hands. Across all language groups, participants consistently provided gesture sequences in which the agent was first, then the patient, and finally the action. By contrast, when participants were asked to verbally describe the same scenes, they used either SVO or SOV according to the basic word order in their native language. How exactly these gestures relate to or consist of linguistic structure is not possible to say; it seems unlikely that participants were producing the grammatical categories Subject-Object-Verb, and perhaps they were not even producing a (single) sentence. Goldin-Meadow et al. (2008) suggest that conveying the agent and patient together before the action is a natural way of representing this kind of event–it highlights the entities involved in an action before the relational action itself. By extension, we can say that in natural language this corresponds to SOV.

This result generated a number of follow-up studies probing the conditions under which gesturers might switch to SVO (Schouwstra & de Swart, 2014; Hall et al., 2014; Marno et al., 2015), and how gesture order might be affected by (or conditioned on) event semantics (Schouwstra & de Swart, 2014; Gibson et al., 2013; Hall et al., 2013), or speaker perspective (Kirton et al., to appear). One of the most intriguing findings from these studies is the apparent difference between gesture order for reversible and non-reversible events. In Goldin-Meadow et al. (2008), almost all the transitive events involved a human actor and an inanimate object patient. These are non-reversible events, with no ambiguity concerning which participant is the agent and which is the patient. Gibson et al. (2013) showed that reversible events–where either participant could in principle play either grammatical role–are more likely to trigger SVO gesture order. For example, when participants were shown a scene with a fireman kicking a ball, they were likely to gesture the fireman first, then the ball, and finally the kicking action. By contrast, when participants were shown a scene with a fireman kicking a girl, they were more likely to gesture the fireman, then the kicking action, then the girl. Futrell et al. (2015a) replicated this pattern of SOV for non-reversible, and SVO for reversible events to a larger set of language populations. This included speakers of two SVO languages (English and Russian), and two VSO languages (Irish and Tagalog), suggesting that the strong subject-initial bias found in previous studies may hold even if gesturers' native language does not conform to this.

Why exactly gesturers shift from SOV to SVO for certain types of events is not yet clear. Gibson et al. (2013) argue that SVO is more robust to noise, preserving information about grammatical relations even if one of the noun phrases is obscured (the 'noisy-channel' hypothesis). Hall et al. (2013, 2014) suggest a number of other possibilities including a gesture-specific constraint on producing a human patient gesture between a human agent and the action it takes. The role of modality-specific constraints is in this case particularly important given the findings reported above using spoken artificial language stimuli. However the differences found between SVO and SOV may stem from constraints on semantic or conceptual naturalness, brought out most clearly in tasks which involve improvisation rather than learning (Schouwstra et al., 2016).

2.5.2 Universal 20

The idea that some word orders might more naturally reflect semantic relations is also at play in the noun phrase. Above we discussed the possibility that Greenberg's Universal 18 might result from a (simplicity) preference for harmony combined with a (naturalness) preference for post-nominal adjectives. Here we will focus on naturalness in the context of Greenberg's 1963 Universal 20, which concerns the relative order of nominal modifiers including adjective, numeral, and demonstrative (rather than how they are ordered relative to the noun). Universal 20 has generated sustained interest among typologists and syntacticians, and the general restrictions posited to be relevant for this universal have been argued to govern word order patterns in other domains as well (e.g., ordering of verbal elements, Koopman & Szabolcsi, 2000). The original statement of the universal highlighted three patterns, picking out Dem-Num-Adj-N as the most common pre-nominal pattern, and N-Dem-Num-Adj along with N-Adj-Num-Dem as the two most common post-nominal patterns. Subsequent work has focused largely on accommodating additional patterns, including many which were unattested in Greenberg's original sample (e.g., Hawkins, 1979; Cinque, 2005; Abels & Neeleman, 2012; Dryer, 2018; Steedman, 2020). Most of these accounts assume semantic or structural distinctions among nominal modifiers that can be described in terms of scope, or semantic compositionality. Intuitively, adjectives modify dimensions inherent to noun meaning and are therefore typically claimed to take innermost scope, composing with the noun first. Numerals then compose with the semantic constituent including both the noun and adjective, therefore taking scope over this unit. Demonstratives serve to connect nominal material to the surrounding discourse, and therefore compose last, taking highest scope over the semantic constituents containing the numeral, adjective and noun (see also Partee, 1987; Adger, 2003; Rijkhoff, 2004). Importantly, these structural relations are argued to influence but not fully determine linear order. Some noun phrase word orders can be derived from it directly, simply by choosing an order for each sub-constituent—N relative to Adj; Num relative to the constituent containing N and Adj; Dem relative to the sub-constituent containing the other two constituents.

There are eight such linearizations, labelled 'homomorphic' following Martin et al. (2020). For example, [Dem [Num [Adj N]]], [[[N Adj] Num] Dem], and [Dem [[N Adj] Num]] are homomorphic (as illustrated by the structurally bracketing), but N-Dem-Num-Adj is not. The homomorphic orders are well attested, and indeed they are all among the most frequent. By contrast, non-homomorphic patterns are generally less common (though based on the latest typological data almost all possible orders are attested, Dryer, 2018). This can be seen most clearly for Greenberg's original three patterns: homomorphic Dem-Num-Adj-N and N-Adj-Num-Dem are the two most frequently attested patterns (note they are also harmonic), while non-homomorphic N-Dem-Num-Adj is attested but rare. Despite decades of work implicating naturalness—i.e., the relation between semantic or conceptual structure and linear order—in this domain, there has been scant behavioral evidence for a bias favoring homomorphic orders.

Culbertson & Adger (2014) provided the first such evidence using the poverty-of-the-stimulus (extrapolation) paradigm. Notably, as mentioned above, this paradigm shares with the silent gesture paradigm the focus on improvising or extrapolating beyond any input data. In Culbertson & Adger (2014), English-speaking adults (Dem-Num-Adj-N) were exposed only to phrases with a single post-nominal modifier, so they had no evidence about the relative order of modifiers. They found that participants nevertheless implicitly assumed an homomorphic order when extrapolating to a phrase with multiple modifiers—i.e., they chose orders like N-Adj-Dem over N-Dem-Adj. These findings were replicated for Thai-speaking adults (N-Adj-Num-Dem) in Martin et al. (2019). Importantly, these studies used real lexical items (either English or Thai) and visually presented phrases to participants. In order to be sure that these results were not driven by a strategy to 'flip' native language orders (thus deriving homomorphic orders without a homomorphism bias), Martin et al. (2020) replicated this study with a fully artificial language, no visual presentation of phrases, and oral production at test.

As for harmony, however, the homomorphism bias found in English and Thai speakers could in principle reflect abstract transfer. The best evidence against this interpretation would be to replicate these results in speakers whose native language is non-homomorphic. That presents a challenge as such languages are rare, therefore Culbertson et al. (2020b) instead used the silent gesture paradigm to investigate whether the bias for homomorphism was present when participants are using a modality distinct from their own. Recall that in work on basic word order using this paradigm, participants did not use their native language order, suggesting that this paradigm may be less likely to elicit native language transfer. Culbertson et al. (2020b) found that English-speaking adults using gesture to convey pictures of simple objects with a pattern or size, numeral, and location contrast consistently improvised homomorphic orders. Moreoever, they showed a bias for post-nominal adjectives, relating back to the typological findings discussed in section 2.4. They also sketch a explanation for how the underlying structure in this domain—which groups adjectives closest to noun, then numerals, with demonstrative furthest away—might be learned by observing objects and their properties in the world. Briefly, they show that an information-theoretic measure of conceptual closeness predicts just this asymmetry among properties conveyed by the three modifier types.

2.5.3 Universal 39

The idea that elements which are conceptual closer to each other is also potentially at play in yet another of Greenberg's 1963 observations: Universal 39. This states that in languages with distinct number marking case marking, number is always placed closer to the noun stem than case. As for the noun phrase, it has been hypothesized that this results from scope or semantic composition; for example, morphemes which more directly affect or modify the semantic content of the stem have narrower scope (e.g. Baker, 1985; Bybee, 1985). However, there are a number of competing hypotheses for why particular patterns of morpheme order might hold, some of which also apply to word order more generally. For example, Hay (2001) argues that morpheme order depends on the degree to which a particular morpheme is easily parseable from the stem (related to how frequently the stem occurs without that morpheme), and relatedly Hahn et al. (2020b,a) argue that word and morpheme order reflect the degree of dependency between the elements in question (related to the information-theoretic notion of surprisal). Both of these accounts are frequency-driven, and do not themselves depend on the meaning of the elements in question. Saldana et al. (2021) test these alternative hypotheses using the extrapolation (Poverty-of-the-Stimulus) paradigm. They trained adult English- and Japanese- speaking adults on a language in which number and case markers were either both pre-nominal or both post-nominal, but there was not information about their relative order. They controlled for frequency, such that nouns were equally likely to occur with number or case marking. Both populations were highly like to infer an order in which number came closer to the noun than case. Evidence from these two populations allows Saldana et al. (2021) to conclude that this finding is likely to be independent from prior language experience; English does not have case, and Japanese does not have number marking (at least not of the type used in the study), and yet both populations showed the same preferences. This suggests that naturalness is indeed at play in morpheme order, although it likely interacts with other surface features of the language, like relative frequencies of different morphemes.[2]

2.6 Communicative Efficiency

Hahn et al. (2020b,a) appeal to the notion of efficiency in their work on word and morpheme order. While the results of Culbertson & Adger (2014) Saldana et al. (2021) suggest this may not necessarily be the ultimate explanation at least for Universals 20 and 39, efficiency has been argued to be at play in a wide range of syntactic (and semantic) domains. In this section, we turn to a set of experiments which aims to test the basic hypothesis that communicative efficiency shapes (morpho)syntax. One common informal definition of efficiency is a balance between simplicity and ambiguity avoidance. A number of studies, often based on cross-linguistic corpora, suggest that languages tend to be efficient—as simple as they can be while maintaining communicative utility (e.g., Gibson et al., 2019; Hahn et al., 2020b). Artificial language learning experiments using a fifth paradigm—typically called the 'iterated learning paradigm'—have shown that languages evolve to be efficient when participants use them to communicate, and then pass them on to new 'generations' of participant-learners (Kirby et al., 2015; Motamedi et al., 2019). However, there is also evidence using the regularization paradigm suggesting that learning on its own may lead to the creation of efficient languages.

2.6.1 Differential case marking

Fedzechkina et al. (2012) explore so-called Differential Case Marking (DCM) systems, a prime example of a phenomenon used to argue for the role of communicative efficiency in language (e.g., Comrie, 1978; Croft, 1990; Jäger, 2007). DCM systems can be characterized as efficient because rather than case marking all event participants (which would minimize ambiguity but maximize complexity), they typically mark only participants that are potentially ambiguous. For example, a typical DCM system might target unusual, or non-prototypical arguments—e.g., inanimate subjects,

[2] This is not the only study suggesting that naturalness may play a role in morphology or morphosyntax. Recent work has also used artificial language learning (both extrapolation and ease-of-learning) to show that certain typologically common person systems (i.e., as instantiated in personal pronoun paradigms) are easier to learn that others, after controlling for paradigm simplicity Maldonado & Culbertson (in press). Maldonado et al. (2020) also use a design similar to Saldana et al. (2021) to explore the relative order of person and number morphemes. In both cases, substantial theoretical work exists, but typological data is sparse, making experimental data of this sort particularly important.

or animate objects. These types of arguments occur in contexts that are potentially ambiguous: if a sentence has an animate object, this object might be mistaken for the subject, if other reliable cues to grammatical roles are absent. The hypothesis that efficiency drives DCM systems has remained in play in the face of typological and theoretical evidence calling it into question (e.g., Aissen, 1999; Haspelmath, 2008; Bickel et al., 2015; Levshina, in press). Fedzechkina et al. (2012) use the regularization paradigm to test whether English-speaking adults will reorganize a language with random variation in word order and case marking such that a DCM system emerges. Participants were taught a language in which case marking appeared randomly on either objects or subjects regardless of their animacy. In the variable object marking condition, by the end of training (which lasted 4 days), participants who produced variable case marking conditioned that marking on animacy, using the marker more often for animate objects. The results were similar, though weaker, for the variable subject case condition. While these results are consistent with an efficiency account of DCM, it is worth noting that participants did not use this language for communication, they simply learned and reproduced it (i.e., described pictures). Moreover, the input languages in this experiment were set up such that the sentences were not functionally ambiguous. Particular nouns in the language were only ever seen as subjects or objects. Thus this experiment does not necessarily provide clear evidence for a communicative efficiency account of case marking. It could instead reflect, for example, a bias to mark unusual alignments between e.g., animacy and grammatical role (e.g., as in Aissen, 1999; Haspelmath, 2008). Smith & Culbertson (submitted) conducted a replication in which the sentences were actually ambiguous, and found that participants indeed created DCM-like systems, but only when they had to actively use the language they had learned to communicate (i.e., describe pictures to a confederate who interpreted them).

2.7 Experiments on processing and perception of (morpho)syntax

2.7.1 Cue position

It is worth nothing that there was also an effect of word order Fedzechkina et al. (2012); case marking was used more often when the case-marked grammatical role was sentence-initial. In the object marking condition, this meant using the marker when the sentence order was OSV, while in the subject marking condition this meant using the marker in SOV sentences (see also Fedzechkina et al., 2017). This is consistent with findings from Pozzan & Trueswell (2015), who show that adult learners are delayed in both comprehension and production of morphology in an artificial language learning task when morphology appears in sentence-final position (Fedzechkina et al., 2015, see also). Pozzan & Trueswell (2015) taught participants one of four languages, which differed along two dimensions. Each language used either head marking or argument marking, and each was either verb-initial or verb-final. In this case, the markers themselves provided information about the type of event (e.g., causative). After two days of learning, on a third day participants were tested on comprehension and production. Performance on both measures was higher for the two verb-initial conditions compared to the two verb-final conditions. The potential role of language processing and comprehension in shaping syntax has been hypothesized by many linguists on the basis of observations from individual languages, and typology (e.g., Hawkins, 2007; Jäger, 2007; Trueswell et al., 2012). These experimental results provide direct behavioral evidence of individual cognitive processes active during learning that can explain structural properties of syntax—here common patterns of case marking, and contingencies between word order and morphology.

2.7.2 Dependency-length minimization

One of the most well-known hypothesized effects of processing on typology is *dependency-length minimization* (DML), the idea that longer dependencies are difficult to process and are generally avoided where possible (e.g., Grodner & Gibson, 2005). This can happen in real time in a given language, when variation is conditioned on phrase length. For example, if we assume that a ditransitive verb like ”took” in the sentence ”Jon took the trash out” has three dependents: ”Jon”, ”out” and ”trash”, then we can shorten the length of the dependencies if we use the alternative ”Jon took out the trash”. Indeed, English exhibits a general short-before-long preference which can be understood in terms of DLM (Arnold et al., 2000). DLM can also happen diachronically, if orders with shorter dependencies become grammaticalized or fixed more often than those with longer dependencies (e.g., Hawkins, 1990). In recent work, Futrell et al. (2015b) show that across a large cross-linguistic sample, languages appear to minimize dependency lengths more than expected by chance. Fedzechkina et al. (2018) test whether English-speaking learners will reorganize an artificial language with variable word order in line with DLM in the lab. The input language featured variable order of subjects and object (SO or OS), and consistent case marking. In one condition, learners were taught a verb-initial language, in the other they were taught a verb-final language. The languages had adjectives which could modify nouns, and adpositional phrases. In the verb-initial language, the latter were prepositional, while in the verb-final language they were postpositional. This meant that shorter dependency lengths could be achieved in the verb-initial language by ordering the subject first and the object second; for example, SO order as in ”[punch] $_{S}$[girl] $_{O}$[boy on red stool]” has shorter dependency lengths than OS order as in ”[punch] $_{O}$[boy on red stool] $_{S}$[girl]”, since the latter case the longer constituent intervenes between the verb and its other dependent. The ordering required to achieve shorter dependencies is reversed in the verb-final language: here, OS order results in shorter dependencies, e.g., ”$_{O}$[boy on red stool] $_{S}$[girl] [punch] ” than SO order as in ”$_{S}$[girl] $_{O}$[boy on red stool] [punch]”. Fedzechkina et al. (2018) find that, indeed, learners take advantage of the variation present in the input language to shorten dependencies in both conditions, using SO order more in the verb-initial condition, and OS order more in the verb-final condition. Importantly, the latter result cannot be explained by the short-before-long preference preference that is already present in English. This study therefore supports the claim that DLM is a bias active during language learning and/or use that may shape languages both synchronically and diachronically. However, as in previous studies reviewed here (e.g., Culbertson et al., 2012; Culbertson & Adger, 2014) there is a possibility of transfer on a more abstract-level: English-speakers have experience with a language in which dependencies are actively minimized, and therefore they may transfer this expectation to a newly learned language. To tackle this possibility, experiments with different populations are again needed.

2.7.3 Affix ordering

As discussed in section 2.5.3, there are good reasons to believe that morpheme and word order are shaped by similar pressures: for naturalness, simplicity, and efficiency, among other things. The kinds of processing-related pressures discussed in the previous section have also been argued to shape features both at the level of syntax and morphosyntax. One well-known example is the so-called ‘suffixing preference’. It has long been noted that more languages use predominantly or exclusively suffixes compared to prefixes (Greenberg, 1963; Dryer, 2013c). Hawkins & Cutler (1988) argue that the preference for suffixes arises due to a confluence of pressures on language processing at the word level. Briefly, they argue that the processing system privileges lexical information over grammatical information, thus it follows that stems should be placed at the beginning and affixes at the end. Hupp et al. (2009) propose that placing important information, like stems at the beginning, and grouping information, like affixes, at the end may reflect a more general perceptual bias. In particular, they show that when English-speaking adults see or hear artificial sequential stimuli (syllables, shapes, musical notes), they rate sequences which differ at their ends as more similar than sequences which differ at their beginnings (see Bruening et al., 2012, for similar findings with English-speaking children). St. Clair et al. (2009) further show that English-speaking adults are better at acquiring two novel categories of artificial lexical items when the categories are marked by an artificial suffix-like element rather than a prefix-like element. This result is also consistent with the idea that humans might have a universal perceptual bias that makes grouping similar words together easier when the similarity is at the end. However, both these studies (and indeed most previous psycholinguistic evidence looking at affix order) test a participant population whose language is heavily suffixing. By contrast, Martin & Culbertson (2020) replicated Hupp et al. (2009) comparing English speakers with speakers of the Bantu language Kîîtharaka, which is predominantly prefixing. While they find that English speakers judge sequences as more similar when they differ at the ends, Kîîtharaka speakers show the opposite pattern, judging sequences that differ at the beginning as more similar. This suggests the possibility that what has been called the 'suffixing preference' in fact does not reflect a cognitive or perceptual bias, but perhaps is a residue of common historical changes (e.g., Himmelmann, 2014).

3 The future of artificial language learning in syntax

The experiments discussed in this chapter represent the current state-of-the-art in using artificial language experiments to investigate key features of syntax (and morphosyntax). The major goal of these studies is to provide empirical evidence connecting recurring patterns of syntactic structure with properties of the human cognitive system. The phenomena I have covered here include nominal and basic word order, morpheme order, case marking and its interaction with word order, and dependency length. I have highlighted along with these phenomena the high-level cognitive and linguistic mechanisms which have been argued to be at play for these phenomena, including simplicity, naturalness, communicative efficiency, and processing ease. In this last section, I would like to discuss three issues which I believe are critical for the future of artificial language learning as a source of evidence for linguistics in general, and syntax in particular.

The first, and perhaps most obvious issue is the range of phenomena studied. While this range is growing year-on-year, it remains narrow in several respects. First, much of the work targets ordering phenomena, which of course does not fully represent (morpho)syntax. Second, for some syntacticians, these phenomena may not be seen as 'core'; there has long been debate about to what degree, for example, Greenberg's Universals are indeed something syntax as a field should seek to explain (e.g., Newmeyer, 2005; Boeckx, 2009). In part this is due to the perception that only exceptionless syntactic generalization should be explained by syntax. But regardless of whether that view has merit, to have wide value to the field, these methods must be applied to a wider range of topics.

In many of the studies discussed here, the range of *participant populations* studied also remains narrow. A full understanding of the extent to which a hypothesized bias is found across speakers of different languages, and across stages of development is crucial in a number of respects. The strongest evidence for the universality of a bias (i.e., evidence of that the bias is at work in all humans, though *not* in all human languages) comes from showing that it is at work even in languages which themselves violate it. Several studies discussed above attempt to provide this kind of evidence, for example in studying the harmony bias in speakers of a non-harmonic languages (Culbertson et al., 2020a), or the subject-first bias in the gesture orders of speakers of a VSO language (Futrell et al., 2015a), or the suffixing-preference with speakers of a prefixing language (Martin & Culbertson, 2020). Further, understanding the role that experience with a particular language, and development more generally, plays in shaping biases will help to build more precise theories of the link between individual cognition and language typology. Whether language structure is shaped primarily by first language learners, by adult second language learners, or by adult patterns of usage continues to be much debated (Yang, 2000; Trudgill, 2011; Lupyan & Dale, 2010). Artificial language experiments have the potential to advance these debates.

Research using artificial language learning experiments must also continue to feed back into theories of grammar. The question of whether and how to explain violable cognitive or linguistic biases is one of the major issues of contemporaries syntax. A number of researchers have attempted to shed light on how syntactic biases can be accounted for in existing theories of grammar (e.g., Culbertson & Smolensky, 2012; Culbertson et al., 2013; Pater, 2011; Perfors et al., 2011). At the same time, some biases may reflect truly domain-general features of cognition, which interact with grammar, but are not directly encoded in the grammatical system itself (e.g., the so-called regularization bias, Reali & Griffiths, 2009; Culbertson et al., 2013; Culbertson & Kirby, 2016). While such biases do not necessarily concern syntacticians directly, a complete account of natural language syntax calls for integrated theories of grammar-internal and external cognitive factors. Indeed, recent computational models have shown that weak biases of these kind are both more likely to have evolved, and able to exert relatively strong effects over time (Thompson et al., 2016). Theories and models of interacting pressures shaping language will help to address a common criticism of work on universal biases in linguistics in general: why don't all languages (eventually) come to confirm to a given bias? The obvious answer is that these biases are defeasible, and thus will interact with competing pressures which potentially pull in other directions. However, artificial language learning paradigms can help to concretize this idea. For example, recent work on the cultural transmission of language has explored the concrete effects of competing pressures for simplicity and ambiguity avoidance using traditional artificial language learning and silent gesture paradigms combined with iterated learning (Kirby et al., 2015; Motamedi et al., 2019).

To summarize, artificial language learning methods are an important method in the toolbox of linguistics for advancing key hypotheses about why language looks the way it does. These methods have been used extensively in work on language development and theoretical phonology, and have recently been extended to syntax, where they have been used to explore the extent to which properties of human cognition are linked to recurring features of syntactic typology. Important directions for these methods in the future include extension to a wider range of hypothesized biases, more comprehensive cross-linguistic and developmental experimentation, and integration with theories of grammar-internal and external forces shaping syntax.

References

Abels, K., & Neeleman, A. (2012). Linear asymmetries and the LCA. *Syntax*, *15*, 25–74.

Adger, D. (2003). *Core Syntax*. Oxford: Oxford University Press.

Aissen, J. (1999). Markedness and subject choice in optimality theory. *Natural Language & Linguistic Theory*, *17*, 673–711.

Ambridge, B., Rowland, C. F., & Pine, J. M. (2008). Is structure dependence an innate constraint? New experimental evidence from children's complex-question production. *Cognitive Science*, *32*, 222–255.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Aristar, A. R. (1991). On diachronic sources and synchronic pattern: An investigation into the origin of linguistic universals. *Language*, *67*, 1–33.

Arnold, J., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, (pp. 28–55).

Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry*, *16*, 373–415.

Baker, M. (2001). *The atoms of language: The mind's hidden rules of grammar*. New York, NY: Basic Books.

Baker, M. C. (2009). Language universals: Abstract but not mythological. *Behavioral and Brain Sciences*, *32*, 448–449.

Biberauer, T., Holmberg, A., & Roberts, I. (2014). A syntactic universal and its consequences. *Linguistic Inquiry*, *45*, 169–225.

Bickel, B. (2007). Typology in the 21st century: Major current developments. *Linguistic Typology*, *11*, 239–251.

Bickel, B., Witzlack-Makarevich, A., & Zakharko, T. (2015). Typological evidence against universal effects of referential scales on case alignment. *Scales and Hierarchies: A Cross-Disciplinary Perspective*, *277*, 7.

Blevins, J. (2004). *Evolutionary Phonology: the emergence of sound patterns*. New York: Cambridge University Press.

Boeckx, C. (2009). Round table—language universals: Yesterday, today and tomorrow. In M. Piattelli-Palmarini, J. Uriagereka, & P. Salaburu (Eds.), *Of Minds and Language : A Dialogue with Noam Chomsky in the Basque Country* (pp. 195–199). Oxford: Oxford University Press.

Braine, M. D. (1963). On learning the grammatical order of words. *Psychological Review*, *70*, 323.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, *29*, 591–610.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, *32*, 76 – 95.

Bruening, P., Brooks, P., Alfieri, L., Kempe, V., & Dabasinskiene, I. (2012). Children's tolerance of word-form variation. *Child Development Research*, .

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711–733.

Bybee, J. (2009). Language universals and usage-based theory. In M. H. Christiansen, C. Collins, & S. Edelman (Eds.), *Language Universals* (pp. 17–39). Oxford: Oxford University Press.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia, PA: John Benjamins.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22.

Chomsky, N. (1957). *Syntactic Structures*. Mouton de Gruyter.

Chomsky, N. (1959). A review of bf skinner's verbal behavior. *Language*, *35*, 26–58.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*, 1–22.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. *Syntax: An international handbook of contemporary research*, *1*, 506–569.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, *36*, 315–332.

Cinque, G. (to appear). A micro-parametric approach to the head-initial/head-final parameter. *Linguistic Analysis*, *42*.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, *20*, 287–333.

Comrie, B. (1978). Ergativity. In W. Lehmann (Ed.), *Syntactic typology: Studies in the phenomenology of language* (pp. 329–394). Austin, TX: University of Texas Press.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, (pp. 522–543).

Croft, W. (1990). *Typology and Universals*. Cambridge Textbooks in Linguistics. New York: Cambridge University Press.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*, 5842–5847.

Culbertson, J., Franck, J., Braquet, G., Barrera Navarro, M., & Arnon, I. (2020a). A learning bias for word order harmony: evidence from speakers of non-harmonic languages. *Cognition*, *204*.

Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonology and semantics in noun class learning. *Journal of Memory and Language*, *92*, 343–358.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain general biases have domain specific effects. *Frontiers in Psychology*, *6*.

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82.

Culbertson, J., & Newport, E. L. (2017). Innovation of word order harmony across development. *Open Mind: Discoveries in Cognitive Science*, *1*, 91–100.

Culbertson, J., Schouwstra, M., & Kirby, S. (2020b). From the world to word order: deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*.

Culbertson, J., & Schuler, K. (2019). Artificial language learning in children. *Annual Review of Linguistics*, *5*, 353–373.

Culbertson, J., & Smolensky, P. (2012). A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science*, *36*, 1468–1498.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*, 306–329.

Culbertson, J., Smolensky, P., & Wilson, C. (2013). Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in Cognitive Science*, *5*, 392–424.

Cysouw, M. A. (2005). Quantitative methods in typology. In *Quantitative Linguistics: An International Handbook* (pp. 554–578). Berlin: Mouton de Gruyter.

DeGraff, M. (1999). Creolization, language change and language acquisition: An epilogue. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony, and development* (pp. 473–543). Cambridge, MA: MIT Press.

Demiral, Ş. B., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from turkish. *Cognition*, *106*, 484–500.

Dryer, M. (1992). The Greenbergian word order correlations. *Language*, *68*, 81–183.

Dryer, M. (2013a). Order of adjective and noun. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dryer, M. (2013b). Order of numeral and noun. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dryer, M. (2018). On the order of demonstrative, numeral, adjective and noun. *Language*, *94*, 798–833.

Dryer, M. S. (1997). On the six-way word order typology. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *21*, 69–103.

Dryer, M. S. (2013c). Prefixing vs. suffixing in inflectional morphology. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dunn, M., Greenhill, S., Levinson, S., & Gray, R. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*, 79–82.

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

Fedzechkina, M., Chu, B., & Florian Jaeger, T. (2018). Human information processing shapes language change. *Psychological science*, *29*, 72–82.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*, 17897–17902.

Fedzechkina, M., Jaeger, T. F., & Trueswell, J. C. (2015). Production is biased to provide informative cues early: Evidence from miniature artificial languages. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 674–679). Austin, TX: Cognitive Science Society.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive science*, *41*, 416–446.

Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, *184*, 53–68.

Fischer, S. D. (1975). *Influences on word-order change in American Sign Language*. Salk Institute for Biological Studies.

Fitch, T., & Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, .

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*, 107–125.

Frigo, L., & McDonald, J. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, *39*, 218–245.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015a). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215–221.

Futrell, R., Mahowald, K., & Gibson, E. (2015b). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*, 10336–10341.

Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, *108*, 17290–17295.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in cognitive sciences*, .

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*, (p. 0956797612463705).

Givón, T. (1979). *On Understanding Grammar*. New York: Academic Press.

Goldberg, A. E. (2013). Substantive learning bias or an effect of familiarity? Comment on Culbertson, Smolensky, and Legendre (2012). *Cognition*, *127*, 420–426.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*, 9163–9168.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of Language* (pp. 73–113). Cambridge, MA: MIT Press.

Grimshaw, J. (1997). Projection, heads, and optimality. *Linguistic Inquiry*, *28*, 373–422.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, *29*, 261–290.

Hahn, M., Degen, J., & Futrell, R. (2020a). Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. Preprint.

Hahn, M., Jurafsky, D., & Futrell, R. (2020b). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, *117*, 2347–2353.

Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating constituent order change with elicited pantomime: A functional account of svo emergence. *Cognitive science*, *38*, 943–972.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*, 1–17.

Haspelmath, M. (2008). Creating economical morphosyntactic patterns in language change. In J. Good (Ed.), *Language universals and language change* (pp. 185–214). Oxford: Oxford University Press.

Hawkins, J. A. (1979). Implicational universals as predictors of word order change. *Language*, *55*, 618–648.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry*, *21*, 223–261.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (2004). *Complexity and efficiency in grammars*. Oxford: Oxford University Press.

Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. *New Ideas in Psychology*, *25*, 87–107.

Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 280–317). Oxford: Blackwell.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, *39*, 1041–1070.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*, 379–440.

Hayes, B. P. (1999). Phonetically driven phonology: The role of optimality theory and inductive grounding. In M. Darnell, E. Moravcsik, F. J. Newmeyer, M. Noonan, & K. Wheatley (Eds.), *Functionalism and formalism in linguistics, vol. 1: General papers*. Amsterdam: John Benjamins.

Himmelmann, N. P. (2014). Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language*, *90*, 927–960.

Hudson Kam, C., & Chang, A. (2009). Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 815.

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation. *Language Learning and Development*, *1*, 151–195.

Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*, 30–66.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*, 876–909.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, *15*, 281–319.

Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, *83*, 74–109.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57*, 129 – 191.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Kirton, F., Schouwstra, M., Culbertson, J., Smith, K., & Kirby, S. (to appear). Constituent order in silent gesture reflects the perspective of the producer. *Journal of Language Evolution*, .

Koopman, H. J., & Szabolcsi, A. (2000). *Verbal complexes*. Cambridge, MA: MIT Press.

Ladd, D. R., Roberts, S. G., & Dediu, D. (2014). Correlational studies in typological and historical linguistics. *Annual Review of Linguistics*, *1*.

Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, *118*, 265–273.

Legate, J. A., & Yang, C. D. (2002). Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review*, *18*, 151–162.

Levshina, N. (in press). Communicative efficiency and differential case marking: A reverse-engineering approach. *Linguistics Vanguard*, .

Lightfoot, D. (1997). Catastrophic change and learning theory. *Lingua*, *100*, 171–192.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*, e8559.

Maldonado, M., & Culbertson, J. (in press). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, .

Maldonado, M., Saldana, C., & Culbertson, J. (2020). Learning biases in person-number linearization. In M. Asatryan, Y. Song, & A. Whitmal (Eds.), *Proceeding of of the fiftieth annual meeting of the North East Linguistic Society* (pp. 163–177). MIT Press volume 2.

Marno, H., Langus, A., Omidbeigi, M., Asaadi, S., Seyed-Allaei, S., & Nespor, M. (2015). A new perspective on word order preferences: the availability of a lexicon triggers the use of SVO word order. *Frontiers in psychology*, *6*.

Martin, A., Abels, K., Ratitamkul, T., & Culbertson, J. (2019). Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard*, *5*.

Martin, A., & Culbertson, J. (2020). Revisiting the suffixing preference: Native language affixation patterns influence perception of sequences. *Psychological Science*, *31*, 1107–1116.

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experiment evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa*, *5*, 1–21.

Martin, A., & White, J. (2019). Vowel harmony and disharmony are not equivalent in learning. *Linguistic Inquiry*, (pp. 1–20).

Martinet, A. (1968). *La linguistique synchronique, Études et recherches*. Paris: Presses Universitaires de France.

Mathews, R. C., & Roussel, L. G. (1997). Abstractness of implicit knowledge: A cognitive evolutionary perspective. In D. C. Berry (Ed.), *How Implicit is Implicit Learning* (pp. 13–47). Oxford: Oxford University Press.

Maurits, L., & Griffiths, T. L. (2014). Tracing the roots of syntax with bayesian phylogenetics. *Proceedings of the National Academy of Sciences*, .

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393–424.

Moeser, S. D., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior*, *11*, 759–769.

Moeser, S. D., & Bregman, A. S. (1973). Imagery and language acquisition. *Journal of Verbal Learning and Verbal Behavior*, *12*, 91–98.

Moreton, E. (2008). Analytic bias as a factor in phonological typology. In C. B. Chang, & H. J. Haynie (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics* (pp. 393–401). Somerville, MA: Cascadilla Proceedings Project.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass*, *6*, 686–701.

Morgan, J. L., Bonamo, K. M., & Travis, L. L. (1995). Negative evidence on negative evidence. *Developmental Psychology*, *31*, 180–197.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*, 498–550.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language*, *28*, 360–374.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of verbal learning and verbal behavior*, *20*, 67–85.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: from improvised gesture to systematic sign. *Cognition*, *192*, 1039642.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Buchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature neuroscience*, *6*, 774–781.

Newmeyer, F. (2005). *Possible and Probable Languages*. New York: Oxford University Press.

Ohala, J. J. (1992). What's cognitive, what's not, in sound change. In G. Kellermann, & M. Morrissey (Eds.), *Diachrony within synchrony: Language history and cognition. Duisberger Arbeiten zur Sprach- und Kulturwissenschaft 14* (pp. 309–355). Frankfurt: Peter Lang.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). London: Longman.

Partee, B. H. (1987). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Dordrecht: Foris.

Pater, J. (2011). Emergent systemic simplicity (and complexity). *McGill Working Papers in Linguistics*, *22*.

Perfors, A., & Burns, N. (2010). Adult language learners under cognitive load do not over-regularize like children. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*. Cognitive Science Society.

Perfors, A., Tenenbaum, J., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*, 306–338.

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

Piantadosi, S. T., & Gibson, E. (2014). Quantitative standards for absolute linguistic universals. *Cognitive Science*, *38*, 736–756.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pozzan, L., & Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive psychology*, *80*, 73–108.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. New York, NY. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Rutgers Optimality Archive 537, 2002. Revised version published by Blackwell 2004.

Pullum, G., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *19*, 9–50.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In M. Tsujimura, & G. Garding (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 101–114). Somerville, MA: Cascadilla Press.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–328.

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

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive science*, *22*, 425–469.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive psychology*, *66*, 30–54.

Rijkhoff, J. (2004). *The Noun Phrase*. Oxford: Oxford University Press.

Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2020). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*, .

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

Saldana, C., Oseki, Y., & Culbertson, J. (2021). Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language*, *118*, 104204.

Saldana, C., Smith, K., Kirby, S., & Culbertson, J. (in press). Is regularisation uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order. *Language Learning and Development*, .

Sandler, W., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *PNAS*, *102*, 2661–2665.

Schouwstra, M., Smith, K., & Kirby, S. (2016). From natural order to convention in silent gesture. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.), *The Evolution of Language*.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*, 431 – 436.

Singler, J. (2006). Children and creole genesis. *Journal of Pidgin and Creole Languages*, *21*, 157–173.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*, 370–407.

Smith, K., & Culbertson, J. (submitted). Communicative pressures shape language during communication (not learning): Evidence from casemarking in artificial languages. Preprint.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 44–449.

Smith, N. V., Tsimpli, I.-M., & Ouhalla, J. (1993). Learning the impossible: The acquisition of possible and impossible languages by a polyglot savant. *Lingua*, *91*, 279 – 347.

St. Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, *33*, 1317–1329.

Steedman, M. (2020). A formal universal of natural language grammar. *Language*, *96*, 618–660.

Steriade, D. (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms.

Tabullo, Á., Arismendi, M., Wainselboim, A., Primero, G., Vernis, S., Segura, E., Zanutto, S., & Yorio, A. (2012). On the learnability of frequent and infrequent word orders: An artificial language learning study. *The Quarterly Journal of Experimental Psychology*, *65*, 1848–1863.

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*, 4530–4535.

Tily, H., Frank, M., & Jaeger, T. (2011). The learnability of constructed languages reflects typological patterns. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1364–1369). Austin, TX: Cognitive Science Society.

Travis, L. (1984). *Parameters and Effects of Word Order Variation*. Ph.D. dissertation MIT.

Trudgill, P. (2011). *Sociolinguistic typology*. New York: Oxford University Press.

Trueswell, J. C., Kaufman, D., Hafri, A., & Lidz, J. (2012). Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada. In *Proceedings of the 36th annual Boston University conference on language development* (pp. 620–632). Somerville, MA: Cascadilla Press.

Vennemann, T. (1976). Categorial grammar and the order of meaningful elements. *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, *3*, 615–634.

de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition*, *107*, 763–774.

Wang, F., Kirby, S., & Culbertson, J. (submitted). Typology reflects learning biases in cross-category harmony. Proceedings of the 43rd Annual Meeting of the Cognitive Science Society.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*, 96–115.

White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a p-map bias. *Language*, *93*, 1–36.

Whitman, J. (2008). The classification of constituent order generalizations and diachronic explanation. In J. Good (Ed.), *Linguistic Universals and Language Change* (pp. 233–252). Oxford: Oxford University Press.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269.

Wilson, C. (2003). Experimental investigation of phonological naturalness. In G. Garding, & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 101–114). Somerville, MA: Cascadilla Press.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982.

Yang, C. D. (2000). Internal and external forces in language change. *Language variation and change*, *12*, 231–250.