THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Simplicity and specificity in language

1 **Simplicity and Specificity in Language:**

2 **Domain general biases have domain specific effects**

3

4 **Jennifer Culbertson* and Simon Kirby**

5 *Language Evolution and Computation Research Unit*

6 *University of Edinburgh, United Kingdom*

7

8 *Correspondence*:*

9 *Jennifer Culbertson*

10 *University of Edinburgh, Dugald Stewart Building, Edinburgh EH8 9AD, United*

11 *Kingdom, jennifer.culbertson@ed.ac.uk*

12

13 **ABSTRACT**

14 The extent to which the linguistic system—its architecture, the representations it operates on,

15 the constraints it is subject to—is specific to language has broad implications for cognitive

16 science and its relation to evolutionary biology. Importantly, a given property of the linguistic

17 system can be "specific" to the domain of language in several ways. For example, if the property

18 evolved by natural selection under the pressure of the linguistic function it serves then the

19 property is domain--specific in the sense that its design is tailored for language. Equally though,

20 if that property evolved to serve a different function or if that property is domain-general, it may

21 nevertheless interact with the linguistic system in a way that is unique. This gives a second

22 sense in which a property can be thought of as specific to language. An evolutionary approach

23 to the language faculty might at first blush appear to favor domain--specificity in the first sense,

24 with individual properties of the language faculty being specifically *linguistic* adaptations.

25 However, we argue that interactions between learning, culture and biological evolution mean

26 any domain-specific adaptations that evolve will take the form of weak biases rather than hard

27 constraints. Turning to the latter sense of domain-specificity, we highlight a very general bias,

28 simplicity, which operates widely in cognition and yet interacts with linguistic representations in

29 domain--specific ways.

30

31 **keywords: language evolution, domain-specificity, simplicity, typological universals,**

32 **compositionality, word order, regularization**

33

34

35

1 INTRODUCTION

One of the fundamental issues in cognitive science is the extent to which specifically linguistic mechanisms and representations underpin our knowledge of language and the way it is learned. This is in part because this issue has deep implications for the underlying uniqueness of a system we typically consider exclusive to humans. It has also been highly divisive in the sense that researchers from distinct traditions often have polar starting assumptions as to the likelihood of domain-specific properties of the language system. Here we will suggest that there are in fact (at least) two ways in which a given feature of the linguistic system may be considered to have domain-specific properties:

(1) If that feature evolved by natural selection under the pressure of the linguistic function it serves.
(2) If that feature is domain-general but interacts with the linguistic system and its representations in a way that is unique.

These two types of domain-specificity are quite different in terms of their implications for the evolution of language, and below we will discuss a set of results from computational models suggesting that domain specificity of the first kind is unlikely to take the form of hard constraints on the linguistic system. Rather, if such constraints exist, they are likely to be weak biases, amplified through cultural evolution. This has important implications for linguistic theory, since, as we discuss below, many mainstream frameworks explicit argue for hard domain-specific constraints and reject the notion of weak bias. The second type of domain-specificity, on the other hand, is likely to be widespread, and highlights the importance of collaborative efforts between experts in linguistic theory–who study the architecture and representations of language–and experts studying cognition across domains and species.

2 DOMAIN SPECIFICITY AND EVOLUTION

In this section, we focus on the first sense of domain-specificity set out above, which interprets the issue in functional terms. This is perhaps the most obvious sense in which a particular aspect of the cognitive system might be specific to language, and it is the one which places a heavier burden on biological evolution. Importantly, it is the ultimate rather than proximate function that is relevant here; knowing that some feature of the cognitive system is used in processing or acquiring language is not, in and of itself, an argument for domain-specificity. We can no more argue that such a feature is language specific because it is active in language

73 processing than we can argue for an aspect of cognition being chess-specific simply because it
74 is active in the brain of a chess player. Rather, we need to consider the *ultimate* function of the
75 cognitive architecture in question by looking to its evolutionary history. An aspect of our
76 cognitive architecture is specific to language if it arose as an adaptive response to the problem
77 of learning or using language.[1]

78   This argument places evolution right at the core of the question of the existence of
79 language-specific features of our cognitive architecture. While some cross-species comparative
80 data exist to help us trace the functional sources of various cognitive capacities (see Fitch, 2010
81 for review), these data are limited by the degree to which the relevant aspects of language are
82 autapomorphies (completely novel traits that are not found in any other species). Recent
83 research has turned to computational modelling to provide a more direct testing ground for
84 specific hypotheses about how the capacities involved in language may have evolved. In
85 particular, a number of papers have looked at whether domain-specific hard constraints on
86 language can evolve from a prior stage where biases were less strong or not present at all (e.g.,
87 Kirby & Hurford, 1997; Briscoe, 2000; Chater, Reali & Christiansen, 2008; Smith & Kirby, 2008;
88 Thompson, 2015). This is important, since many linguistic theories conceive of the language
89 capacity as including a set of constraints of this kind: for example, Biberauer, Holmberg &
90 Roberts (2014), working in the Minimalist framework (Chomsky, 1993), argue for a constraint
91 which places a hard (inviolable) restriction on the distribution of the feature triggering movement
92 (they call it the 'Final-Over-Final' constraint, in a nod to the structural description of word orders
93 the constraint rules out). Similarly, in Optimality Theory (Prince & Smolensky, 1993/2004),
94 although a particular constraint may be violated in a given language, the standard mechanism
95 for explaining typological data is to restrict the set of constraints. For example Culbertson,
96 Smolensky & Wilson (2013) describe an OT grammar for word order in the noun phrase which
97 completely rules out particular patterns by using a limited set of so-called alignment constraints
98 (see also Steddy & Samek-Lodovici, 2011).

99   To investigate how hard domain-specific constraints of this type might evolve, Chater,
100 Reali & Christiansen (2009) describe a simulation of a population of language-learning agents.
101 The genes of these agents specify whether learning of different aspects of language is tightly
102 constrained or highly flexible. Agents in the simulation that successfully communicate are more
103 likely to pass on their genes to future generations. The question that Chater et al., (2009) ask is
104 whether genes encoding constraints evolve in populations which start out highly flexible under
105 the selection pressure for communication. If they do, then this would support a language faculty
106 in which language acquisition is constrained by domain-specific principles. This process,

---

[1] Note this is true even if we then happen to use this aspect of our cognitive system for other, additional purposes. The fact that we use our language faculty for solving crosswords does not constitute an argument against domain specificity of that faculty.

107 whereby traits that were previously acquired through experience become nativised, is known as
108 the *Baldwin Effect* (Baldwin, 1896; Maynard Smith, 1987; Hinton & Nowlan, 1987), and a
109 number of authors have suggested it played a role in the evolution of the language faculty
110 (Turkel, 2002; Kirby & Hurford, 1997; Jackendoff, 2002). However, Chater et al., (2009) argue
111 that the fact that languages change over time makes the situation of language evolution quite
112 different from that of other learned traits. In their simulations, if the rate of language change is
113 high enough, it is impossible for genetic evolution to keep up–language presents a moving
114 target, and domain-specific constraints cannot evolve.
115   Chater et al.'s (2009) model is a critique of a particular view of the language faculty in
116 which hard innate constraints are placed on the form languages can take. Because of this they
117 do not model a scenario in which the strength of bias is allowed to evolve freely (although they
118 do show that their model gives similar results whether genes encode hard constraints, or very
119 strong biases). However, there is growing support for a more nuanced view of language
120 acquisition in which learners have biases that come in a range of strengths (e.g., Morgan, Meier
121 & Newport, 1989; Wilson, 2006; Hudson Kam & Newport, 2009; Smith & Wonnacott, 2010;
122 Culbertson & Smolensky, 2012; Culbertson et al., 2013; Chater, Clark, Goldsmith & Perfors,
123 2015). If the genes underpinning the language faculty were able to specify everything from a
124 very weak bias all the way to a hard constraint, then perhaps this would allow evolution to take a
125 gradual path from an unbiased learner to a strongly-constraining, domain-specific language
126 faculty. To find out if this is the case, we need a model that shows how bias strength affects the
127 nature of the languages that emerge in a population.
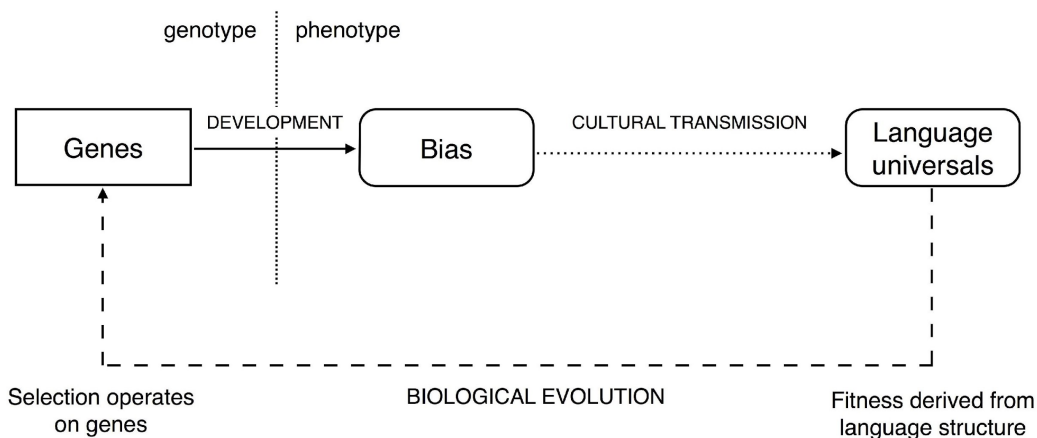
128
129 **Figure 1.** The link between genes and the universal properties of language is mediated by
130 development and cultural transmission. The extent to which these two processes have non-
131 trivial dynamics is an important consideration when proposing evolutionary accounts of
132 language. Fitness does not depend directly on the genes underpinning the language faculty, but
133 rather the linguistic phenotype (i.e. languages). This opens up the possibility for development

134 and cultural transmission to shield genetic variation from the view of natural selection. (Figure
135 adapted from Kirby et al, 2007).
136
137   The iterated learning model (Kirby, Dowman & Griffiths, 2007) starts from the
138 observation that the way languages evolve culturally is driven by the way in which languages
139 are learned.[2] This model of cultural evolution suggests that the languages spoken by a
140 population will not necessarily directly reflect the learning biases of that population (Figure 1). In
141 particular, in many cases, cultural evolution will tend to amplify weak learning biases. This has
142 important implications for how constraints on the language faculty actually come to be reflected
143 in properties of language. For example, the observation that some property of language is
144 universally, or near universally, present in language is not sufficient for us to infer that there is a
145 corresponding strong constraint in our language faculty. Indeed, if Kirby et al., (2007) are
146 correct, then the strength of any constraint in the language faculty may be *unrelated* to the
147 strength of reflection of that constraint cross-linguistically. Weak learning biases may be
148 sufficient to give rise to exceptionless, or near exceptionless, universals.
149   Smith & Kirby (2008) examine the implications of iterated learning for the biological
150 evolution of the language faculty. Their simulation explicitly models three processes involved in
151 the origins of linguistic structure: individual learning of languages from data; cultural evolution of
152 languages in a population through iterated learning; and biological evolution of learning biases
153 themselves. They show that neither hard constraints nor strong biases emerge from the
154 evolutionary process even when agents are being selected for their ability to communicate using
155 a shared language. This is a consequence of the amplifying effect of cultural evolution; the
156 fitness of an organism is not derived directly from that organism's genes, but rather from the
157 organism's phenotype. In the case of language evolution this is the actual language an
158 individual has learned. If weak learning biases are amplified by cultural evolution, then the
159 difference between a weak bias and a hard constraint is neutralised: both can lead to strong
160 effects on the distribution of languages. What this means is that iterated learning effectively
161 masks the genes underpinning the language faculty from the view of natural selection. They are
162 free to drift; strongly-constraining domain-specific constraints on language learning are likely to

---

[2] Our emphasis in this article will be on learning, but there are other mechanisms that operate at the individual level but whose effect is felt at the population level. For example, the way in which hearers process input, and the way in which speakers produce output is likely to have a significant impact. See Kirby (1999) for an extended treatment of precisely how processing and learning interact with cultural transmission to give rise to language universals, and Futrell et al., (2015), Fedzechkina et al., (2012), and Jaeger & Tily (2011) for recent accounts of specific links between processing and language structure. However, the debate about domain generality/specificity plays out differently for processing than for learning, and as such will not be the focus of this review. In particular, here we discuss *simplicity* as a highly general learning bias that unifies a range of different domains both within and beyond language, and it is not clear that an equivalent notion of simplicity exists for processing.

163   be lost due to mutation, or not arise in the first place (see also, Thompson, 2015 for a detailed

164   analysis of the evolutionary dynamics in this case).

165   Taken together these modelling results show that domain-specific hard constraints on

166   language learning are unlikely to evolve, because languages change too fast (Chater et al.,

167   2009) and because cultural evolution amplifies the effect of weak biases (Kirby, Dowman &

168   Griffiths, 2007). However, the results of this latter model suggest a further conclusion: weak

169   biases for language learning are *more* evolvable by virtue of cultural evolution's amplifying

170   effect. Any tiny change from neutrality in learning can lead to big changes in the language that

171   the population uses. Just as culture masks the strength of bias from the view of natural

172   selection, it unmasks non-neutrality. We argue that linguists should not shy away from

173   formulating domain-specific aspects of the language faculty in terms of weak, defeasible biases.

174   This is the type of language faculty that is most likely to evolve.

175   Although we propose that strong domain-specific biases on language should be avoided

176   on evolutionary grounds, this does not mean that strong domain-*general* biases are impossible.

177   These may be the result of very general architectural or computational considerations that

178   govern the way cognition operates, for example (falling under the third of Chomsky's (2005)

179   three factors in language design). Equally, the way we learn language might be shaped by

180   relatively strong domain-general biases that arise as a result of evolution for something other

181   than language, for which the amplifying effect of culture does not apply. Biases such as these

182   may nevertheless interact with language and linguistic representations in domain-specific ways.

183   In the next section we will examine a learning bias that is arguably the most domain-general of

184   all–simplicity–and show how its application in a range of different aspects of language leads to

185   domain-specific outcomes.

186

187   **3 SIMPLICITY**

188

189   Simplicity has been proposed as a unifying principle of cognitive science (Chater & Vitanyi,

190   2003). The tradition of arguing for a general simplicity bias has a long history in the context of

191   scientific reasoning dating back to William of Occam in the 14th century who stated that we

192   should prefer the simplest explanation for some phenomenon all other things being equal. In

193   other words, when choosing among hypotheses that explain data equally well, the simpler one

194   should be chosen.

195   This principle can be extended straightforwardly from scientific reasoning to cognitive

196   systems. When faced with an induction problem we must have some way of dealing with the

197   fact that there are many candidate hypotheses that are consistent with the observed data

198   (typically an infinite number). So, for example, in a function learning task how do we interpolate

199   from seen to unseen points when there are an infinite number of possible functions that could

200 relate the two (Figure 2)? Or, to give a more trivial example, why is it that we assume that the
201 sun will continue to rise every day when there are an infinite range of hypotheses available to
202 us which predict it won't.
203
204
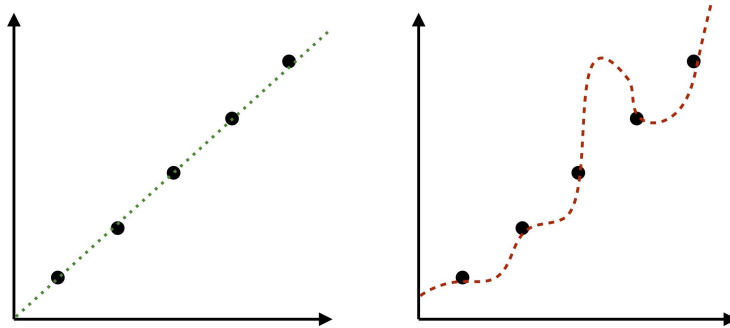
Figure 2. There are an infinite set of possible functions interpolating from seen points to unseen
points in these graphs. Our intuition is that the linear function on the left represents a more
reasonable hypothesis than the one on the right, despite the fact that both fit the data perfectly
well. In other words, we have prior expectations about what functions are more likely than
others. In this case, the prior includes a preference for linearity (cf. Kalish, Griffiths &
Lewandowsky, 2007).

213 Here again the simplicity bias provides an answer by giving us a way to distinguish between
214 otherwise equally explanatory hypotheses. While a full treatment of why simplicity rather than
215 some other bias is the correct way to solve this problem is beyond the scope of this article
216 (accessible introductions are given in Mitchell, 1997; Chater et al., 2015), we can give an
217 intuitive flavour in terms of Bayesian inference. According to Bayes rule, induction involves
218 combining the probability distribution over hypotheses defined by the data with a *prior*
219 probability distribution over these hypotheses. More formally, the best hypothesis, , for some
220 data, D, will maximize $P(D|H)P(H)$.

$$ = \quad {}_{\in} \ (\ |\ ) = \quad {}_{\in} \ (\ |\ )(\ ) $$

222 What can this tell us about simplicity? We can express this equivalently by taking logs of these
223 probabilities. The best hypothesis is the one that *minimises* the sum of negative log probabilities
224 of the data given that hypothesis, $-\ _2\ (\ |\ )$, and the prior probability of the hypothesis
225 itself, $-\ _2\ (\ )$.

$$ = \quad {}_{\in} \ -\ {}_2\ (\ |\ ) = \quad {}_{\in} \ -\ {}_2\ (\ |\ ) -\ {}_2\ (\ ) $$

226

7

227 Information theory (Shannon & Weaver, 1948) tells us that this last quantity, $-\log_2(\ )$, is the

228 description length of in bits (assuming an optimal encoding scheme for our space of

229 hypotheses). So, all other things being equal, learners will choose hypotheses that can be

230 described more concisely–hypotheses that are simpler.

231 Importantly, an information theoretic view of the equation above also suggests learners

232 will prefer representations that provide (to a greater or lesser extent) some compression of the

233 data they have seen. What does this mean for the nature of language? It suggests that

234 languages will be more prevalent to the extent that they are compressible. In general, a

235 language will be compressible if there are patterns within the set of sentences of that language

236 that can be captured by a grammatical description. More precisely, a compressible set of

237 sentences is one whose minimum description length is short. The description length is simply

238 the sum of the length of the grammar ($-log_2 P(h)$ in the equation above) and the length of the

239 data when described using that grammar (given by the $-log_2 P(D|h)$ term).

240 This argument has allowed us to relate our intuitive understanding of simplicity–as a

241 reasonable heuristic in choosing between explanations–to a rational model of statistical

242 inference in a relatively straightforward way. Of course, there are a lot of practical questions that

243 this leaves unanswered. How, for example, can we tell in a given domain what counts as a

244 simpler hypothesis? Unfortunately, there is no computable general measure of complexity (Li &

245 Vitanyi, 1997), nevertheless we propose that notions of relative simplicity should guide our

246 search for domain general biases underpinning phenomena of interest in language.

247 So, we argue that–whatever other biases learners have when they face some learning problem–

248 they are also likely to be applying an overarching simplicity bias (Chomsky, 1957; Clark, 2001;

249 Brighton, 2002; Kemp & Regier, 2012; Chater et al., 2015).

250 It is important to note that when we talk about simplicity in the context of language, it is

251 in terms of the overall compressibility of that language, e.g. how much redundancy and

252 systematicity does it exhibit that can be captured simply in a grammatical description, and how

253 much irreducible unpredictability remains in the data. We might also be interested in ways in

254 which languages differ in the length of their utterances, but this is a largely orthogonal issue.

255 Indeed, it is possible for a language with shorter strings to have a longer grammar–consider

256 cases of irregular morphology in which regularization might simplify a paradigm at the cost of

257 removal of short irregulars.

258 The generality of the bias for simplicity suggests there will be many linguistic

259 phenomena affected by it. Below, we discuss cases which have been documented both in

260 linguistic typological *and* experimental studies, with an emphasis on morphology and syntax (for

261 discussion of experimental findings related to phonological simplicity, see Moreton & Pater

262 2012a,b). We will begin with a basic design feature of language–compositionality–that can be

263 characterized by the interaction of simplicity with a competing pressures for expressivity. We
264 then move on to three additional examples of increasingly narrow phenomena: regularization of
265 unconditioned variation, consistent head ordering or word order harmony, and isomorphic
266 mapping from semantic structure to linear order. Each example illustrates a slightly different
267 way in which this domain-general bias interacts with features that are particular to the linguistic
268 domain.
269

270 ### 3.1   Compositionality
271 For our first example we will consider a basic property of language, often called a "design
272 feature" (Hockett, 1960): the compositional nature of the mapping between meanings and
273 forms. Language is arguably unique among naturally occurring communication systems in
274 consisting of utterances whose meaning is a function of the meaning of its sub-parts and the
275 way they are put together. For example, the meaning of the word 'stars' is derived from the
276 meaning of the root *star* combined with the meaning of the plural morpheme *-s*. Similarly, the
277 meaning of a larger unit like 'visible stars' is a function of the meanings of the individual parts of
278 the phrase. Switching the order to 'stars visible' changes the meaning of the unit in a predictable
279 way.[3]

280   This ubiquitous feature of language makes it arguably unique among naturally occurring
281 communication systems, the vast majority–perhaps all–of which are *holistic* rather than
282 compositional (Smith & Kirby, 2012). The striking divergence from holism that we see in
283 language (above the level of the word) is therefore of great interest to those studying the
284 evolution of language. The fact that human communication is also highly unusual in consisting
285 of learned rather than innate mappings between meanings and signals suggests that relating
286 the origins of compositionality to learning biases is a good place to start in the search for an
287 explanation.

288   A language that maps meanings onto signals randomly (see Figure 3A) will be less
289 compressible–and hence, less *simple* in our terms–than one which maps them onto signals in a
290 predictable way (see Figure 3B). Where both signals and meanings have internal, recombinable
291 structure, then this predictability will be realised as compositional mappings. To see why this is,
292 consider representing language as a *transducer* relating meanings and signals. The transducer
293 in Figure 4A gives the most concise representation of an example holistic language, whereas

---

[3] In this case, placing the adjective after the noun leads to the interpretation "the stars visible
(tonight)". This is a systematic rule of English: post-nominal attributive adjectives are stage-level
predicates, denoting temporary properties (Cinque, 1993).

294   the transducer in Figure 4B gives the most concise representation of an equivalent

295   compositional language in which subparts of the signals map onto subparts of the meanings.

296   What should be immediately apparent is that compositional languages are more compressible.

297
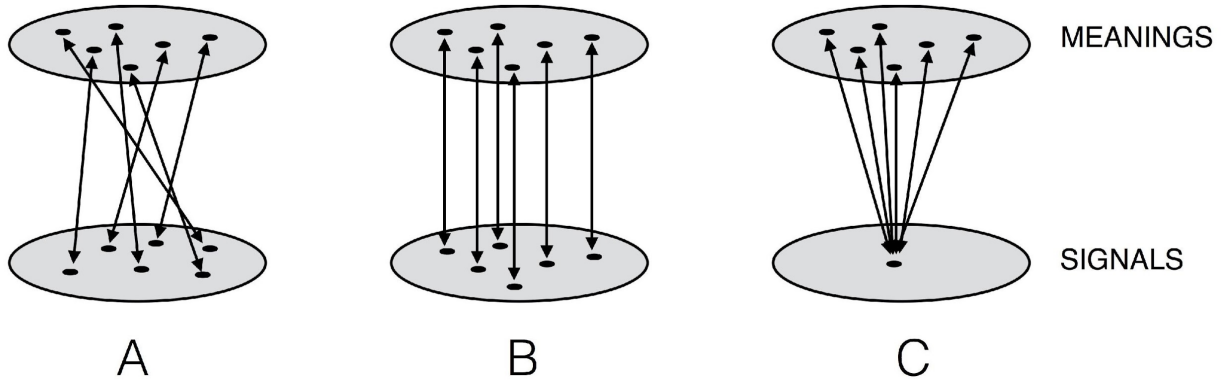
298

299   **Figure 3**. A simplified geometric sketch of possible mappings between two domains, for

300   example meanings and signals. These mappings can be unstructured, random and

301   incompressible (A), or highly structured and compressible (B). An individual attempting to learn

302   the latter could use similarity structure in one domain to predict what the appropriate

303   generalisation should be for unseen points. A further possibility is a degenerate mapping, which

304   is the simplest and most compressible of all (C).

305

A.	S : GO PAST	→ went	**B.**	S : x y	→ A:x B:y
	S : BE PAST	→ was		A : MOVE	→ move
	S : LIE PAST	→ lay		A : EXIST	→ exist
	S : TEACH PAST	→ taught		A : REST	→ rest
	S : GO PRES	→ go		A : LEARN	→ learn
	S : BE PRES	→ am		B : PAST	→ ed
	S : LIE PRES	→ lie		B : PRES	→ ∅
	S : TEACH PRES	→ teach			

306

307   **Figure 4**. Two simple transducers that map between a subset of the English verbs and their

308   meanings, where "S" is the start symbol for the transducers and meanings are given in caps

309   after a colon in each rule. Transducers can be *holistic*, essentially a dictionary of meaning-signal

310   pairs (A); or *compositional*, in which the meaning of a signal is composed of the meaning of

311   parts of that signal (B).

312

313   Brighton (2002) uses this contrast to model the cultural evolution of compositionality in
314   an iterated learning framework (Kirby et al., 2007). Individual agents in their simulation learn
315   transducers to map between a structured set of meanings and signals made up of sequences of
316   elements. Crucially, the learners have a prior bias in favour of simpler transducers. In fact, the
317   prior probability of a particular transducer is inversely related to its coding length in bits in
318   precisely the way outlined in our discussion of simplicity above. Each agent learns their
319   language by observing meaning-signal pairs produced by the previous agent in the simulation,
320   and then goes on to produce meaning-signal pairs for transmission to the next generation. As
321   the language in these simulations is repeatedly learned and reproduced, the bias of the agents
322   in favour of simplicity shapes the evolutionary dynamic. Despite the fact that these models
323   involve no biological evolution, the grammars adapt gradually over cultural generations from
324   ones that are random and holistic to ones that are compositional.[4]
325   This result makes intuitive sense if you think about the process of transmission from the
326   point of view of the emerging rules and regularities in the mapping between meanings and
327   signals. A highly specific feature of the evolving language (e.g. a particular idiosyncratic label for
328   a single meaning, like *went* as the past tense of GO) will be harder to learn than a
329   generalisation over a large number of meanings (e.g. a morpheme, like *–ed*, that shows up in
330   the signals associated with a wide range of meanings). Particularly if learners only see a subset
331   of all possible meanings, this inevitably leads to a preferential transmission of broader and
332   broader generalisations that apply across large parts of the language. Hurford (2000) puts it
333   pithily, stating "social transmission favours linguistic generalisation".
334   The simplicity bias thus appears to predict one of the fundamental design features of
335   human language. However, things are not quite so straightforward. Consider a language in
336   which every meaning is expressed by the same signal (Figure 3C). This *degenerate* language
337   will be even more compressible than the compositional one, suggesting that a domain-general
338   bias for simplicity is not sufficient to explain the origins of compositional structure. Cornish
339   (2011) argues that in fact all simulations of iterated learning purporting to demonstrate the
340   emergence of compositionality have in some way implemented a constraint that rules out

---

[4] Brighton (2002) makes the simplicity bias of the learners in his model overt by counting the numbers of bits in the encoding of transducers that generate the data the learners see. However, this does not mean that we necessarily believe that this kind of representation of grammars is necessary for an implementational or algorithmic account of what language we are doing when they learn language. Rather, this is a computational level account in Marr's (1982) terms. It is an empirical question whether the particular ranking of grammars in terms of simplicity that we can derive from this particular representation matches precisely the ranking that applies in the case of real language learners, but we are confident that the crucial distinction between degenerate < compositional < holistic is correct. This matches behaviour of participants in the lab (Kirby et al., 2015) and broadly similar results are found in both connectionist and symbolic models of iterated learning (Kirby & Hurford, 2002; Brace, Bullock & Noble, 2015).

341 degeneracy. It is simply impossible for the learners in these simulation models to acquire a

342 language that maps many meanings to one signal. Similarly, in the first laboratory analog of

343 these iterated learning simulations, Kirby, Cornish & Smith (2008) report that degenerate

344 languages rapidly evolve over a few generations of human learners.

345   Kirby, Tamariz, Cornish & Smith (2015) argue that a countervailing pressure for

346 expressivity is required to avoid the collapse of languages in iterated learning experiments to

347 this degenerate end point. The obvious pressure arises not from learning, but from use. If pairs

348 of participants learn an artificial language and then go on to use it in a dyadic interaction task,

349 then there are two pressures on the language in the experiment: a pressure to be compressible

350 arising from participants' domain-general simplicity bias in learning, and a pressure to be

351 expressive arising from participants' use of the language to solve a communicative task. Kirby et

352 al. (2015) show that compositionality only arises when both of these two pressures are in play.

353 In this case then, a domain-general bias is only explanatorily adequate once we take into

354 account features of domain of application. In other words, the case of compositionality illustrates

355 that the simplicity bias is domain-specific in the sense that we cannot understand how it shapes

356 language without also appealing to the special function of language as a system of

357 communication.

358

359 **3.2   Regularization**

360 There is converging evidence from multiple strands of research including pidgin/creole studies,

361 sociolinguistics, language acquisition, and computational cognitive science suggesting that

362 language tends to minimize unpredictable or unconditioned variation. Variation can be

363 introduced by non-native speaker errors, contact with speakers of other languages, or in the

364 case of newly emerging languages, variation may reflect a lack of conventionalized grammar. In

365 the latter case, there is evidence that new generations of learners regularize and

366 conventionalize these noisy systems (e.g., Sankoff 1979; Mühlhäusler 1986; Meyerhoff 2000;

367 Senghas & Coppola 2001). Natural language and laboratory language learning research has

368 further shown that both children and adults learn and reproduce conditioned variation relatively

369 well compared to unpredictable variation (e.g., Singleton & Newport, 2004; Hudson Kam &

370 Newport, 2005, 2009; Smith, Durham & Fortune, 2007; Smith & Wonnacott, 2010; Culbertson,

371 Smolensky & Legendre, 2012). For example, Singleton & Newport (2004) report the case of a

372 child acquiring American Sign Language (ASL) from late-learner parents. While the parents'

373 realization of several grammatical features of ASL was variable, the child did not reproduce this

374 variation. Rather, he regularized his parents' variable productions, resulting in a much more

375 consistent system (though in some aspects it differed from ASL). Following up on this finding

376 using an experimental paradigm, Hudson Kam & Newport (2009) report that, when trained on a

377 grammar with unpredictable use of determiners, child learners (and to a lesser extent adults)

378 regularize those determiners, using them according to a consistent rule.

379   Computational modeling has formalized this in terms of learners' a priori expectations,

380 namely that observed data come from a deterministic generative process (Reali & Griffiths,

381 2009; Culbertson & Smolensky, 2012; Culbertson et al., 2013). This has a natural interpretation

382 in terms of simplicity, since the description of a language that only allows one option in a

383 particular context will be shorter than one that allows multiple variants.[5] More generally, as

384 we've seen already, there's a straightforward relationship between the entropy of the distribution

385 of variants and the coding length of that distribution. More predictable processes can be

386 captured by shorter overall descriptions: they are compressible (Ferdinand, 2015). However, the

387 expectation that the world will be deterministic is to some extent dependent on the domain in

388 question. Most obviously, prior experience in a given domain can override this expectation–e.g.,

389 we expect that a coined being tossed will be fair and therefore outcomes will be random (Reali

390 & Griffiths, 2009). In a carefully controlled study comparing learning of unpredictable variation in

391 a linguistic versus non-linguistic domain, Ferdinand (2015) found that regularization occurs in

392 both domains. However, across a number of conditions manipulating system complexity, the

393 bias is stronger for linguistic stimuli. Regularization thus illustrates a case in which the *strength*

394 of a bias is domain-specific, perhaps dependent on previous experience and functional

395 pressures relevant to that domain.

396   While most recent work on regularization focuses on unconditioned or random variation,

397 there is some evidence that even conditioned variation is avoided in language. For example,

398 English is losing its system of irregular (variable) past tense marking in favor of a single rule

399 (add -*ed*) despite this variation being lexically conditioned (Hooper, 1976). Similarly, while some

400 languages allow widespread lexically or semantically conditioned variation in adjective

401 placement, most languages tend to order them more or less consistently before or after (Dryer,

402 2013). This can be related straightforwardly to simplicity; a grammar with a single (high-level)

403 rule or constraint applying to all words of a given type is more compressible than one in which

404 different such words must obey different rules. For example, a grammar with a single rule

405 stating that adjectives must always precede nouns is simpler than one which has to specify that

406 certain adjectives precede and others follow.

407

408 **3.3   Harmony**

---

[5] Note that this requires taking into account the simplicity of the generating grammar *and* the simplicity (compressibility) of the data. A grammar which allows free variation may be simpler than a grammar which generates conditioned variation, however the random data produced by the former grammar is not compressible.

409 Interestingly, this reflex of simplicity applies not only to word order within a word class, but also
410 across classes of words. Some of the best known typological universals describe correlations
411 among words orders across different phrase types. For example, Greenberg (1963) lists a set of
412 universals, collated from a sample of 30 languages, including the following:
413
414   *Universal 2*: In languages with prepositions, the genitive almost always follows the
415   governing noun, while in languages with postpositions it almost always precedes.
416
417   *Universal 18:* When the descriptive adjective precedes the noun, the demonstrative and
418   the numeral, with overwhelmingly more than chance frequency, do likewise.
419
420 These universals are part of the evidence for word order harmony–the tendency for a certain
421 class of words to appear in a consistent position, either first or last, across different phrase
422 types in a given language (Greenberg, 1963; Chomsky, 1981; Hawkins, 1983; Travis, 1984;
423 Dryer, 1992; Baker, 2001; for experimental evidence see Culbertson et al., 2012; Culbertson &
424 Newport, 2015). At its root, this is just an extension of the same very general statement of
425 within-category order consistency. However, absent a notion of what ties certain categories of
426 words together, the connection between harmony and simplicity remains opaque. For example,
427 the two universals quoted above make reference to a single category–noun–and how it is
428 ordered relative to a number of other categories. Based on syntactic class alone, simplicity
429 predicts that nouns should be ordered consistently relative to all these other categories. This is,
430 of course, the wrong prediction; Universal 2 actually says that the order of nouns relative to
431 adpositions is the *opposite* of the order of nouns relative to genitives. While adpositions and
432 genitives thus tend to appear on different sides of the noun, it turns out that adjectives,
433 demonstratives, and numerals often pattern with genitives (note that English is a
434 counterexample). These tendencies are exemplified in (3).
435
436   3) a. Preposition N {Adj, Num, Dem, Gen}
437   b. {Adj, Num, Dem, Gen} N Postposition
438
439 To make sense of this, we need a notion that connects adpositions as they relate to nouns, with
440 nouns as they relate to the other categories. The most popular such notion provided by linguistic
441 theory is the head-dependent relation. In this example, the noun is a head with respect to
442 nominal modifiers–including genitive phrases, adjectives, numerals, and demonstratives. By
443 contrast, the noun is a dependent in an adpositional construction. When stated in this way,
444 harmony falls out: in the world's languages, there is a tendency for heads to consistently
445 precede *or* follow their dependents. The former type is often called head-initial, the latter head-

14

446 final. Coming back to simplicity then, a language which has a single high-level rule stating that
447 heads either precede or follow their dependents is simpler than one which has specific ordering
448 rules for heads in distinct phrase types. Simplicity therefore predicts that the more specific rules
449 a grammar has, the less likely it should be.

450   Importantly, a clear understanding of whether this prediction is borne out depends on the
451 precise definition of the relevant relation between word categories. This turns out to be
452 controversial. For example, particular theories differ in what is deemed to be a head, and
453 whether 'head' is in fact the relevant notion at all (Hawkins, 1983; Zwicky, 1985; Hudson 1987;
454 Dryer, 1992; Corbett et al., 1993). Dryer (1992) provides typological evidence that head order
455 does not correlate across all phrase types. For example, he reports that the order of verb (head)
456 and object (dependent) correlates with the order of preposition (head) and noun (dependent)
457 within a language, but not with noun (head) and adjective (dependent) order. This is unexpected
458 if the simplicity bias is indeed based on head-dependent order. He therefore argues that a
459 different notion, related to the average length or complexity of particular phrase types, must be
460 used in order to see that languages do indeed prefer higher-level rules governing order across
461 multiple phrase types. Regardless of whether Dryer's precise formulation is correct, what this
462 suggests is that merely stating that simplicity is a factor in determining word order does not
463 allow us to determine which grammars are in fact the simplest. In order to do this, we need a
464 theory of linguistic representations which tells us which should be treated as parallel and in what
465 contexts.

466   From the perspective of the learner, there is also a clear sense in which the simplicity
467 bias as it relates to word order harmony depends on linguistic representations. Given three
468 words, in the absence of any knowledge about the relations between and among them, there is
469 no way simplicity can be used by a learner to make inferences about likely orderings. These
470 representations must be present (e.g., learned) before a simplicity bias can be active. How and
471 when they develop–i.e., when particular syntactic categories are differentiated, when abstract
472 higher-level categories like head develop, etc.–will dictate how simplicity impacts learners'
473 inferences.

474

475 **3.4   Isomorphic mapping**
476 The relation between word order and semantic interpretation in a number of domains also
477 appears to be affected by a simplicity bias. For example, Greenberg's (1983) Universal 18
478 describes how nominal modifiers are ordered relative to the noun. Universal 20 builds on this,
479 describing how those modifiers tend to be ordered relative to one another.

480

481   *Universal 20* (as restated by Cinque, 2005)*:*

482   In pre-nominal position the order of demonstrative, numeral, and adjective (or any

483   subset thereof) is *Dem-Num-Adj*.

484   In post-nominal position the order is either *Dem-Num-Adj* or *Adj-Num-Dem*.

485

486   Interestingly, while both post-nominal orders are indeed possible, addition typological

487 work since Greenberg (1963) indicates that the second order is *much* more common. In fact,

488 Dem-Num-Adj-N, and N-Adj-Num-Dem are the two most common orders found in the world's

489 languages by far. Part of this is likely due to the harmony bias described above; assuming

490 nominal modifiers are covered by the relevant notion of dependent, these two orders are

491 harmonic, while alternative possibilities are not (e.g., Dem-Num-N-Adj). However harmony does

492 not explain why N-Adj-Num-Dem would be more common than N-Dem-Num-Adj. An

493 explanation of this difference depends on how syntax–specifically, linearization–interacts with

494 underlying semantic structure.

495   Several theoretical lines of research converge on a universal semantic representation of

496 these modifiers and their relation to the noun. On one view, this representation reflects iconicity

497 of relations (Rijkhoff, 2004). For example, adjectives modify inherent properties of nouns,

498 numerals count those larger units, and demonstratives connect those countable units to the

499 surrounding discourse. This describes a nesting representation as in Figure 5A. Research in

500 formal linguistics further suggests a hierarchical relation between these elements in terms of

501 semantic combination, illustrated in Figure 5B. Crucially, these abstract relations are preserved

502 in linear orders that have the adjective closest to the noun and the demonstrative most

503 peripheral–orders that can be read directly off Figure 5A. Notice that N-Adj-Num-Dem is one

504 such order, while N-Dem-Num-Adj is not (the modifiers must be swapped around to get this

505 order). Recent laboratory studies suggest a corresponding cognitive bias, in favor of isomorphic

506 mappings between nominal semantics and linear order (Culbertson & Adger, 2014). Typological

507 frequency differences in this domain can be therefore be much better explained once we take

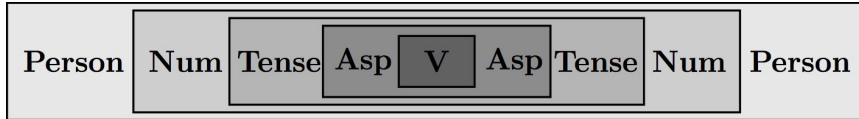508 into account the underlying semantic structure and an isomorphism bias.

509

510

511 A   Dem | Num | **Adj** | **N** | **Adj** | Num | Dem   B
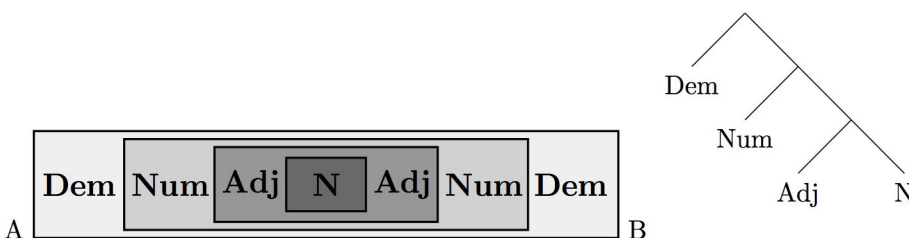
512 **Figure 5**. Nested representation (A) and hierarchical representation (B) of semantic relations

513 between modifiers and the noun. The most typologically common orders can be read off

514 directly.

515

516

517 **Figure 6**. Schematic representation of semantic composition in verbal domain.

518

519   This is not the only case of isomorphic mappings from semantics to linear order, indeed

520 perhaps the most well-known case is the mirror principle in the domain of verbal inflection

521 (Baker, 1985; Bybee, 1985; Rice, 2000). Languages tend to order inflectional morphemes like

522 tense and aspect in a way that reflects semantic composition, as shown in Figure 6.[6]

523   Biases in favor of isomorphism between semantics and linear order can again be

524 reduced to a general simplicity bias. In very general terms, more transparent or predictable

525 relations between order and meaning are simpler than ones with extra arbitrary stipulations.

526 Brighton & Kirby (2006) show that isomorphic[7] mappings between signals and meanings arise

527 naturally from iterated learning under general simplicity considerations. Put in more precise

528 terms, to derive surface order from semantics, each branch of the hierarchical structure (or each

529 rectangle in the nested schematic) in the figure above represents a choice point for linearization.

530 For isomorphic orders, that is all that is required: N-Adj-Num-Dem means choosing (1) Adj after

531 N, (2) Num after [N-Adj], and (3) Dem after [N-Adj-Num]. Similarly, a non-harmonic but

532 isomorphic order like Dem-Num-N-Adj is (1) Adj after N, (2) Num before [N-Adj], and (3) Dem

533 before [Num-N-Adj]. By contrast, non-isomorphic orders require additional choice points or

534 rules. N-Dem-Num-Adj, for example, cannot be derived from the semantic hierarchy alone–the

535 simplest route is Dem-Num-Adj-N (three choice points) plus one addition rule placing N first.

536 The isomorphism bias again illustrates that the notion of simplicity, however general, must be

[6] Interestingly, the acquisition of semantics literature provides a related observation. Musolino,
Crain & Thornton (2000) show that when asked to interpret ambiguous sentences with
quantificational elements, children strongly prefer the interpretation that corresponds to the
surface syntactic position of those elements. For example, the sentence "Every horse didn't
jump over the fence", could involve *every* taking scope over *not* (meaning no horses jumped
over the fence), or *not* scoping over *every* (meaning not every horse jumped over the fence).
The first interpretation is isomorphic to the linear order, and this is the interpretation preferred by
young children (see also Musolino & Lidz, 2003).
[7] These authors use the term "topographic" rather than "isomorphic" because of similarity to the
neuroanatomical organising principle of topographic maps. For our purposes the terms are
interchangeable, since both give rise to the property that neighbouring representations in one
domain map to neighbouring representations in the other.

537   formulated with reference to specific hypotheses about the domain in question–here, about

538   conceptual iconicity or formal compositional semantics.

539

540   **4 CONCLUSION**

541   There is little doubt that the language faculty includes capacities and constraints that are

542   domain-general or co-opted from other cognitive systems. Whether it also includes domain-

543   specific features is both less clear, and more likely to split along philosophical lines; traditionally,

544   generative linguistics has argued for a Universal Grammar containing (among other things)

545   linguistically contentful principles that place hard constraints on what is learnable. We have

546   suggested, based on results obtained using computational models of language evolution, that

547   domain-specific hard constraints are much less likely to have evolved than weak biases. This is

548   essentially because the cultural evolution of language exerts cognition-external pressures that

549   mean linguistic phenotypes no longer directly reflect the underlying genotype. The strength of

550   any particular bias is underdetermined by the cross-linguistic distribution of language types. At

551   the same time, these cognition-external pressures allow weak genetically-encoded biases to

552   have potentially large typological effects. While this does not categorically rule out the existence

553   of very strong (or inviolable) biases that have evolved specifically for language, it clearly

554   suggests we should not treat them as the default hypothesis. The idea that weak biases for

555   language-specific structures or patterns are more likely is in line with recent trends in linguistics.

556   Researchers in phonology and syntax have begun using formal models which encode

557   probabilistic biases in order to better capture empirical data from typology and learning (e.g.,

558   Hayes & Wilson, 2008; Pater, 2009; Culbertson et al., 2013; White, 2014).

559   Regardless of whether the language faculty contains domain-specific capacities, the

560   representations which make up our linguistic knowledge, and the function of language as a

561   system of communication means that domain-*general* capacities will interact with language in

562   unique ways. This is most convincingly illustrated by looking at an uncontroversially general

563   bias: the bias in favor of representational simplicity. The examples we have discussed here

564   show that a simplicity bias is reflected in a range of language universals that cut across very

565   different aspects of the linguistic system: compositionality, regularity, harmony, and

566   isomorphism. In each case, the simplicity bias interacts with linguistic representations to give

567   rise to domain-specific effects. In the case of compositionality, simplicity interacts with the major

568   unique function of language as a communication system that must be expressive. It is only via

569   the interaction of these two pressures that compositional systems will emerge. The

570   regularization bias, which describes the established finding that language learners tend to

571   reduce random or unconditioned variation, shows domain-specific effects in terms of its

572   strength. Word order harmony, the tendency for languages to order heads consistently before or

573   after dependents, depends crucially on a language- and even theory-specific notion of the

574 relevant categories. Finally, the notion of isomorphism between semantic or conceptual
575 structure and surface word order crucially requires an articulated hypothesis about the specific
576 semantic relations among dependent elements.
577   In all these cases, distinct hypotheses about linguistic categories, their representations,
578 and how they relate to one another will make distinct predictions about how simplicity is cashed
579 out. This means that an understanding of language, how it is learned, and how it evolved will
580 necessarily require input from linguists formulating theories of the architecture and
581 representations of language. The fact the many aspects of the capacity for language also come
582 from broader cognition means linguists in turn must take into account findings from research on
583 other cognitive domains, and indeed on related capacities in other species.
584
585
586
587

588 **References**
589
590 Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry*, *16*,
591   373–415.
592 Baker, M. (2001). *The atoms of language: The mind's hidden rules of grammar*. New York, NY:
593   Basic Books.
594 Baldwin, J. M. (1896). A new factor in evolution (continued). *American Naturalist*, (pp. 536–553).
595 Biberauer, T., Holmberg, A., & Roberts, I. (2014). A syntactic universal and its consequences.
596   *Linguistic Inquiry*, *45*, 169–225.
597 Brace, L., Bullock, S., & Noble, J. (2015). Achieving compositional language in a population of
598   iterated learners. In P. Andrews, L. Caves, R. Doursat, S. Hickinbotham, F. Polack, S.
599   Stepney, T. Taylor, & J. Timmis (Eds.), *Advances in Artificial Life, ECAL 2015* (pp. 349–
600   356). MIT Press.
601 Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*, 25–54.
602 Brighton, H., & Kirby, S. (2006). Understanding linguistic evolution by visualizing the emergence
603   of topographic mappings. *Artificial Life*, *12*, 229–242.
604 Briscoe, T. (2000). Grammatical acquisition: Inductive bias and coevolution of language and the
605   language acquisition device. *Language*, (pp. 245–296).
606 Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*.
607   Philadelphia, PA: John Benjamins.
608 Chater, N., Clark, A., Perfors, A., & Goldsmith, J. A. (2015). *Empiricism and Language
609   Learnability*. Oxford University Press.
610 Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in
611   language evolution. *Proceedings of the National Academy of Sciences*, *106*, 1015–1020.

612   Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in*
613   *Cognitive Sciences*, *7*, 19–22.
614   Chomsky, N. (1957). *Syntactic Structures*. Mouton de Gruyter.
615   Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
616   Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale, & S. J. Keyser (Eds.),
617   *The View from Building 20; Essays in Linguistics in Honor of Sylvain Bromberger* (pp. 1–52).
618   Cambridge, MA: MIT Press.
619   Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*, 1–22.
620   Cinque, G. (1993). On the evidence for partial N-movement in the Romance DP. *University of*
621   *Venice Working Papers in Linguistics*, *3*, 21–40.
622   Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, *36*,
623   315–332.
624   Clark, R. (2001). Information theory, complexity, and linguistic descriptions. In S. Bertolo (Ed.),
625   *Parametric Linguistics and Learnability* (pp. 126–171). Cambridge: Cambridge University
626   Press.
627   Corbett, G. G., Fraser, N. M., & McGlashan, S. (1993). *Heads in grammatical theory*.
628   Cambridge: Cambridge University Press.
629   Cornish, H. (2011). *Language adapts: Exploring the cultural dynamics of iterated learning*. Ph.D.
630   thesis, University of Edinburgh.
631   Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over
632   surface frequency. *Proceedings of the National Academy of Sciences*, *111*, 5842–5847.
633   Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of
634   language universals. *Cognition*, *139*, 71–82.
635   Culbertson, J., & Smolensky, P. (2012). A Bayesian model of biases in artificial language
636   learning: The case of a word-order universal. *Cognitive Science*, *36*, 1468–1498.
637   Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order
638   universal. *Cognition*, *122*, 306–329.
639   Culbertson, J., Smolensky, P., & Wilson, C. (2013). Cognitive biases, linguistic universals, and
640   constraint-based grammar learning. *Topics in Cognitive Science*, *5*, 392–424.
641   Dryer, M. (1992). The Greenbergian word order correlations. *Language*, *68*, 81– 183.
642   Dryer, M. (2013). Order of adjective and noun. In M. S. Dryer, & M. Haspelmath (Eds.), *The*
643   *World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary
644   Anthropology.
645   Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their
646   input to facilitate efficient communication. *Proceedings of the National Academy of*
647   *Sciences*, *109*, 17897–17902.

648   Ferdinand, V. (2015). *Inductive evolution: cognition, culture, and regularity in language*. Ph.D.
649   thesis, University of Edinburgh.

650   Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.

651   Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length
652   minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*,
653   10336–10341.

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

657   Hawkins, J. A. (1983). *Word order universals*. New York: Academic Press.

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

660   Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution. *Complex systems*, *1*,
661   495–502.

662   Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88–96.

663   Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological
664   change. In W. Christie (Ed.), *Current progress in historical linguistics* (pp. 96–105).
665   Amsterdam: North-Holland.

666   Hudson, R. A. (1987). Zwicky on heads. *Journal of Linguistics*, *23*, 109–132.

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

669   Hurford, J. R. (2000). Social transmission favours linguistic generalization. In *The evolutionary
670   emergence of language: Social function and the origins of linguistic form* (pp. 324–352).
671   Cambridge: Cambridge University Press.

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

674   Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative
675   efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 323–335.

676   Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational
677   knowledge transmission reveals inductive biases. *Psychonomic Bulletin* & *Review*, *14*, 288–
678   294.

679   Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general
680   communicative principles. *Science*, *336*, 1049–1054.

681   Kirby, S. (1999). *Function, Selection, and Innateness*. Oxford: Oxford University Press.

682   Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An
683   experimental approach to the origins of structure in human language. *Proceedings of the
684   National Academy of Sciences*, *105*, 10681–10686.

685   Kirby, S., Dowman, M., & Griffiths, T. (2007). Innateness and culture in the evolution of
686   language. *Proceedings of the National Academy of Sciences*, *104*, 5241.
687   Kirby, S., & Hurford, J. (1997). Learning, culture and evolution in the origin of linguistic
688   constraints. In *Fourth European Conference on Artificial Life* (pp. 493– 502). Citeseer
689   Cambridge, MA: MIT Press.
690   Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the
691   cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
692   Marr, D. (1982). *Vision: A computational approach*. San Francisco: W. H. Freeman.
693   Meyerhoff, M. (2000). The emergence of creole subject–verb agreement and the licensing of
694   null subjects. *Language Variation and Change*, *12*, 203–230.
695   Ming, L., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*.
696   Heidelburg: Springer Science & Business Media.
697   Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
698   Moreton, E., & Pater, J. (2012a). Structure and substance in artificial-phonology learning, part I:
699   Structure. *Language and Linguistics Compass*, *6*, 686–701.
700   Moreton, E., & Pater, J. (2012b). Structure and substance in artificial-phonology learning, part II:
701   Substance. *Language and Linguistics Compass*, *6*, 702 – 718.
702   Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with
703   cross-sentential cues to phrase structure. *Journal of Memory and Language*, *28*, 360–374.
704   Mühlhäusler, P. (1986). *Pidgin and Creole Linguistics*. Oxford: Blackwell.
705   Musolino, J., Crain, S., & Thornton, R. (2000). Navigating negative quantificational space.
706   *Linguistics*, *38*, 1–32.
707   Musolino, J., & Lidz, J. (2003). The scope of isomorphism: Turning adults into children.
708   *Language Acquisition*, *11*, 277–291.
709   Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science*, *33*, 999–
710   1035.
711   Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative*
712   *grammar*. New York, NY. Technical Report, Rutgers University and University of Colorado at
713   Boulder, 1993. Rutgers Optimality Archive 537, 2002. Revised version published by
714   Blackwell 2004.
715   Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating
716   regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–328.
717   Rice, K. (2000). *Morpheme order and semantic scope: Word formation in the Athapaskan verb*.
718   Cambridge: Cambridge University Press.
719   Rijkhoff, J. (2004). *The Noun Phrase*. Oxford: Oxford University Press.
720   Sankoff, G. (1979). The genesis of a language. In K. C. Hill (Ed.), *The Genesis of Language*
721   (pp. 23–47). Ann Arbor, MI: Karoma Publishers.

722   Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign
723   Language acquired a spatial grammar. *Psychological Science*, *12*, 323–328.

724   Shannon, C. E., & Weaver, W. (2015). *The mathematical theory of communication*. Illinois:
725   University of Illinois Press.

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

728   Smith, J., Durham, M., & Fortune, L. (2007). "mam, my trousers is fa'in doon!": Community,
729   caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation*
730   *and Change*, *19*, 63–99.

731   Smith, J. M. (1986). When learning guides evolution. *Nature*, *329*, 761–762.

732   Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human
733   language faculty and its evolution. *Philosophical Transactions of the Royal Society B:*
734   *Biological Sciences*, *363*, 3591–3603.

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

737   Steddy, S., & Samek-Lodovici, V. (2011). On the ungrammaticality of remnant movement in the
738   derivation of Greenberg's Universal 20. *Linguistic Inquiry*, *42*, 445–469.

739   Thompson, B. (2015). *Transmission, Induction, and Evolution*. Ph.D. thesis, University of
740   Edinburgh.

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

742   Turkel, W. J. (2002). The learning guided evolution of natural language. In T. Briscoe (Ed.),
743   *Linguistic Evolution through Language Acquisition: Formal and Computational Models*
744   chapter 8. Cambridge: Cambridge University Press.

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

747   Wilson, C. (2006). An experimental and computational study of velar palatalization. *Cognitive*
748   *Science*, *30*, 945–982.

749   Zwicky, A. M. (1985). Heads. *Journal of linguistics*, *21*, 1–29.