



## Discussion

## Substantive learning bias or an effect of familiarity? Comment on Culbertson, Smolensky, and Legendre (2012)

Adele E. Goldberg\*



Princeton University, Green Hall, Princeton, NJ 08544, United States

## ARTICLE INFO

## Article history:

Received 28 March 2012

Revised 21 February 2013

Accepted 22 February 2013

Available online 29 March 2013

## Keywords:

Learning bias

Transfer effect

Universal 18

Universal Grammar

Constructions

## ABSTRACT

Typologists have long observed that there are certain distributional patterns that are not evenly distributed among the world's languages. This discussion note revisits a recent experimental investigation of one such intriguing case, so-called "universal 18", by Culbertson, Smolensky, and Legendre (2012). The authors find that adult learners are less likely to generalize an artificial grammar that involves the word order combination Adjective-before-Noun and Noun-before-Numeral, and they attribute this to two factors: (1) a domain-general preference for consistency—i.e., a preference for either N *before* Adj/Num, or N *after*, and (2) a domain-specific unlearned universal bias against Adj-N + N-Num order. An alternative explanation for the second factor is that it involves a transfer effect from either Spanish-type languages or from English. The case for possible transfer from English is based on the fact that adjectives regularly occur after the nouns they modify in several English constructions, whereas numerals only quantify the nouns they follow in one construction that occurs extremely infrequently.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

While generative linguists have traditionally focused on the possible existence of exceptionless universals (e.g., Kayne, 1994), they have begun to take an interest in statistical trends as well (e.g., Prince & Smolensky, 1993/2004; Yang, 2004). There are various possible causes of distributional asymmetries, including shared historical roots and language contact (e.g., Dryer, 2000; Dunn, Greenhill, Levinson, & Gray, 2011), or cognitive-functional factors related to the purpose of language as a communicative system (e.g., Bybee, 2009; Croft, 2003; Goldberg, 2006), and experimental work has recently begun to investigate whether such trends may result from violable learning biases (Ellefsen & Christiansen, 2000). This discussion note focuses on a recent experimental investigation of one such

intriguing case by Culbertson, Smolensky, and Legendre (2012) (hereafter, CSL): so-called "universal 18".

Greenberg (1963) first observed that possible orderings of adjectives and numerals within the noun phrase are not distributed equally. In particular, the fourth word order pattern in Table 1 is relatively rare (the numbers are provided by CSL based on the WALS sample, Dryer, 2008a; Dryer, 2008b).

The four word orders in Table 1 do not exhaust all logical possibilities, and other possibilities are attested as well. Over 100 languages have variable orders of adjective and noun (Dryer, 2008a), with neither order dominant, and over 50 have variable orders of numeral and noun (Dryer, 2008b). There also exist languages that have no numerals (Frank, Everett, Fedorenko, & Gibson, 2008), or no distinct grammatical category of adjective (Dixon, 1977).

While the A-N + N-Num order is rarer than the other orders represented in Table 1, it is the dominant order in two dozen known languages (Dryer, 2000). Dryer cites

\* Tel.: +1 609 258 8772.

E-mail address: [adele@princeton.edu](mailto:adele@princeton.edu)

**Table 1**

Fixed word orders of Adj/Noun and Numeral/Noun.

Ordering	Examples	Languages	% (# of lngs)
1. Adjective-Noun & Numeral-Noun	<i>Red birds &amp; three birds</i>	English, Cherokee	27% (227)
2. Noun-Adjective & Noun-Numeral	<i>Birds red &amp; birds three</i>	Japanese, Yoruba	52% (443)
3. Noun-Adjective & Numeral-Noun	<i>Birds red &amp; three birds</i>	Sp., It, Hebrew	17% (149)
4. Adjective-Noun & Noun-Numeral	<i>Red birds &amp; birds three</i>	Sinhala, Newari	4% (32)

the following example from Purki (note that A-N and N-Num combine to yield A-N-Num):

- (1) rdamo bomo ngis  
 beautiful girl two  
 'two beautiful girls' (Rangan, 1979: 122)

This impressive variability among the world's languages should make one wary of overstating linguistic regularities. Yet, if it turns out that A-N + N-Num order is harder to learn than other fixed orders, it might indicate that there is a reason why this order is rarer typologically. But crucially, the next natural question is, *why* should this order be more difficult to learn?

## 2. Culbertson, Smolensky, Legendre (2012)(CSL)

In their very interesting recent paper, CSL investigate the possible existence of a substantive learning bias against A-N + N-Num word order, where a substantive learning bias is defined as a universal bias to "acquire grammars that do not incorporate a particular [disfavored] structure" (CSL: 308). CSL taught undergraduate participants the meanings and novel labels for 10 nonce nouns, five adjectives, and five numerals. After learning the vocabulary, participants were exposed to combinations of two words, adjective and noun; or numeral and noun.

Four separate groups of undergraduates were exposed to mini-languages in which one of the languages in Table 1 occurred 70% of the time. The alternative order was witnessed the other 30% of the time. The dominant pattern in each of CSL's conditions 1–4 corresponded to the word orders 1–4 in Table 1. For example, condition 1 was the English-like A-N, Num-N grammar, and so the probability of witnessing A-N or Num-N was 70%, and the probability of witnessing the opposite order (N-A or N-Num) was 30%. The order was not fixed for any particular noun or modifier, but was instead generated randomly. After exposure to the mini artificial language, participants took part in a production task. There was incentive to try to anticipate the word orders that would be generated by the probabilistic grammar as points were awarded for correctly matching the order that was generated probabilistically.

The design relied on previous results that have found a domain-general tendency for participants to regularize just this sort of unconditioned, irregular input, in effect simplifying the pattern by reducing variability. For example, this tendency has been found in artificial grammar learning tasks (Hudson Kam & Newport, 2009) and in non-linguistic pattern matching studies (Gardner, 1957; Weir, 1972), particularly when accuracy is rewarded, as it was in the CSL study. Of interest was whether participants would regularize each of the language conditions to the same extent.

CSL's results demonstrate that learners are quite likely to regularize the *consistent* ("harmonic") orders in which both A or Num are ordered either before or after the noun—i.e., the first two orders in Table 1. This regularization bias is domain-general, since as CSL note, it exists throughout cognition and motor learning. *Ceteris paribus*, one thing (e.g., "N comes first") is easier to learn than two (e.g., "N appears before A" and "N appears after Num").

The finding that does not immediately yield to a simple domain-general explanation is that condition 3's N-A + Num-N (*grifta blue & three grifta*) is generalized somewhat more than condition 4's A-N + N-Num (*blue grifta & grifta three*). This is in fact "the comparison of greatest interest" (p. 317), as it represents CSL's proposed substantive bias. The key theoretical claim the authors aim to contribute is that this bias is universal and domain-specific: "If Universal 18's substantive bias against a particular type of non-harmonic language [A-N + N-Num] is in fact specific to the language system, then the empirical findings reported here constitute clear evidence against recent claims that no such biases exist within cognition" (CLS: 326).

In fact, CSL find only a marginal effect of condition in this key comparison, with a significant interaction between condition and modifier type. The interaction was driven by a difference in numerals (cf. Fig. 3, p. 317): participants who witnessed Num-N as the majority order (condition 3) were significantly more likely to use that order than were participants who witnessed N-Num as the majority order likely to use N-Num order (condition 4). Interestingly, there was only a marginally significant difference in how readily participants generalized the N-A (of condition 3) vs. A-N (of condition 4) order. We return to this point below.

## 3. Cognitive biases

There is no question but that cognitive biases of some sort exist. Humans are not born blank slates, and it is clear that humans can learn language while the star-nosed mole cannot. No one denies this. What is at issue is whether there exist learning biases that are unlearned, specific to language, and which do not serve a useful function. Traditional generative syntax has long assumed that these types of substantive biases do exist in the form of a Universal Grammar (Chomsky, 1965), but as CSL observe, there is a growing chorus of challenges to that assumption. Most challenges to the Universal Grammar Hypothesis emphasize the fact that universals (or biases) that are specific to language are more rare than is often assumed, and when they exist they serve some *function* within the language system (Bates, 1993; Beckner et al., 2009; Bybee, 2009;

Christiansen & Chater, 2008; Elman, Bates, et al., 1996; Goldberg, 2004, 2006; Haspelmath, 2008; Simone & Vallauri, 2010; Tomasello, 2003, 2009).

CSL are explicitly agnostic about the “locus, scope, experience–dependence, and ultimate source” (p. 307) of the bias against A-N + N-Num. They are also noncommittal about whether the bias has a functional basis or not; they suggest two possible explanations: one apparently non-functional and one functional, as reviewed in Section 4. They emphasize, however, that the bias “does not plausibly reflect a domain-general constraint: it therefore constitutes evidence for the existence of cognitive biases specific to language” (p. 323). They also assume that such “prior probabilities of possible grammars” exist as part of learners’ “initial knowledge” (p. 320).

The question of whether a bias serves a function is critical to explaining how the bias may come to exist. If a bias serves a clear function, it is not necessary to assume that it is innate or given *a priori*. For example, paraphrasing Liz Bates, the fact that all unimpaired humans eat with their hands and not their feet does not mean that there is an innate bias against eating with one’s feet. Eating with our hands is simply a more efficient (and cleaner) way to get food into our mouths than using our feet. The function itself—its usefulness—provides a reason for the bias to arise during the course of development without appeal to natural selection.

It is worth trying to be explicit about what an unlearned domain-specific functionless constraint against A-N + N-Num would entail. Where would it come from? Can a bias against a possible word order possibly be biologically encoded? How and why would such a specific constraint evolve? Surely it is not life-threatening nor sexually unattractive to produce the A-N + N-Num order. An alternative might be to suggest that it is a *spandrel*; that is, it may be a byproduct of some other feature that does have an evolutionary advantage. But no such account has been suggested. Alternatively, if a non-functional bias is *not* to have arisen through natural selection, we are owed an account of how and why it exists (see Blumberg, 2006; Deák, 2000; Karmiloff-Smith, 1994 for relevant discussion). With this in mind, we turn the two possible explanations for a bias against A-N + N-Num order that CSL propose. It will be argued that ultimately neither provides a compelling explanation of the experimental data. In Section 5, we revisit the possibility rejected by CSL that greater generalization in condition 3 compared to condition 4 is simply a transfer effect.

#### 4. Possible explanations for a bias against A-N + N-Num order

##### 4.1. The final over final constraint (FOFC) proposal

The first possible motivation offered by CSL for the preference for type 3 over type 4 languages is that it may be an instance of a syntactic “final over final constraint” (FOFC) that makes reference to the notion of *head* (Biberauer, Holmberg, & Roberts, 2008). A head of a phrase is the word that determines the character and properties of the phrase

it is a head of (Croft, 1996; Lakoff, 1970; Polinsky, 2012; Zwicky, 1985). The FOFC stipulates that a head-initial phrase  $[\alpha c]_{\alpha P}$  cannot be embedded within a head-final phrase  $[X \beta]_{\beta P}$  (the subscripts “ $\alpha P$ ” and “ $\beta P$ ” are used to indicate that  $\alpha$  and  $\beta$  are the heads of their respective phrases). This constraint rules out the structure:  $*[[\alpha c]_{\alpha P} \beta]_{\beta P}$ .<sup>1</sup>

Note that the only way the FOFC can account for the dispreference of  $[[A N] \text{ Numeral}]$  is if the adjective, not the noun, is considered the head of the  $[A N]$  phrase. Accordingly, CSL offer the structure:  $[[A NP]_{AP} \text{ Numeral}]_{\text{Numeral phrase}}$  (their Fig. 7, p. 326; cf. also Abney, 1987; Cinque, 2005). But the idea that the adjective is the head of A + N combinations is quite problematic. It is the noun that typically determines the character of the adjective–noun combination, not the adjective; e.g., a *red bird* is a bird, not a “red”. Moreover, the adjective is quite often dependent on the noun for its syntactic and semantic properties, including plural and gender agreement. Thus without redefining what it means to be the head of a phrase, appeal to the FOFC constraint does not account for CSL’s intriguing findings.

In addition, in order to allow for the fact that participants in condition 4 produced the ill-formed structure more than 60% of the time (Fig. 2, p. 315), and the fact that at least two dozen languages prefer the supposedly ill-formed A-N-Num order, the FOFC would, as CSL point out, require a reinterpretation as a violable constraint. Finally, since the FOFC constraint appears to be an example of a non-functional bias, it raises the question of how it came to exist.

##### 4.2. A processing proposal based on a head primacy principle

CSL alternatively suggest a possible functional basis for their findings, citing Kamp and Partee (1995)’s head primacy principle (which assumes the N, not the A is the head):

Head primacy principle: In a modifier-head structure, the head  $[N]$  is interpreted relative to the context of the whole constituent [the NP], and the modifier [e.g., A] is interpreted relative to the local context created from the former context by the interpretation of the head  $[N]$  (Kamp & Partee 1995: 161).

The head primacy principle is motivated by the fact that the interpretation of adjectives often depends on the nouns they modify as well as other aspects of context. Clearly a *skillful violinist* is only claimed to be skillful in the context of violin-playing, a *short mountain* is only short as far as mountains go, and so on (this fact provides additional evidence against the idea that the A is the head of the  $[A N]$  phrase, insofar as the adjective is dependent on the N for its interpretation).

CSL interpret the head primacy principle as claiming that nouns must be interpreted before adjectives, regardless of overt word order, which would seem to predict a

<sup>1</sup> The FOFC is a much weakened version of the more familiar “head-direction parameter” that had claimed that all heads must align.

preference for N-A order, although Kamp and Partee do not themselves make any processing claims or claims about preferred linear order. According to CSL's summary based on WALS (Dryer, 2008a,b; Table 1), 69% of languages use N-A order, while 31% use A-N, so it is conceivable that there may be some weak dispreference for A-N order. However, this does not account for CSL's data, since, as CSL acknowledge in note 37 (p. 325), "the differences across conditions are not carried by the adjective phrases, as would be predicted if there is a bias against Adj-N but not against N-Num".<sup>2</sup>

Moreover, there was no uncertainty nor temporary ambiguity in the production task of the experiment. Participants knew that one or more entities of a certain type were represented on the basis of the visual image alone, and they also knew that they needed to use an NP to describe each image due to the experimental design. That is, no disadvantage in interpretation nor parsing was relevant within the experimental context.

To summarize, neither of the proposed explanations suggested by CSL are compelling motivations for the learning bias against A-N + N-Num order that was in evidence in CSL's comparison between participants' tendencies to generalize in condition 3 vs. condition 4. Perhaps a more plebian explanation is worth a second look. Familiarity with Spanish or English may underlie the key comparison.

## 5. A simple familiarity (transfer) effect

The participants in the CSL study were undergraduates at Johns Hopkins and might be expected to be familiar to varying extents with Spanish or other Latinate languages that tend to be type 3 languages (N-A + Num-N), or with Modern Hebrew, another type 3 language. Is this what provides the key advantage of condition 3 over condition 4? There is strong evidence that a second language can affect the learning of a third language, possibly even more than a first language (Falk & Bardel, 2001). To CSL's credit, they attempt to address this, a reviewer's concern, in their Section 5.3.2. The authors sent questionnaires to all of the undergraduates who had participated, in order to learn about exposure to type 3 languages, and significant correlations were not in evidence. Unfortunately, only 23 responses were received, and these were distributed among four conditions. Thus, the results are inconclusive and the

<sup>2</sup> A related type of processing advantage could be suggested on the basis of Hawkins (1994)'s argument that typological tendencies in linear ordering are correlated with the rapidity with which immediate constituents of a phrase can be produced and recognized on-line. However, in the experimental context, every phrase was a two-word noun phrase consisting of a noun and either an adjective or a numeral. Since adjective and numeral never cooccurred, half of the tokens included the numeral-less A-N pattern which requires a one word delay in building an N, and therefore should be dispreferred. And yet condition 1 (along with condition 2) was the most readily generalized. There may well be merit to the head primacy principle and/or to Hawkins' general observations about the order of constituents. They are both based on a domain-general preference to lessen demands on working memory. Insofar as either generalization can motivate the typological facts, they may provide domain-general explanations. But they do not appear to motivate CSL's findings.

possibility of a transfer effect from participants' familiarity with type 3 languages remains.<sup>3</sup>

In addition to the possibility of a transfer effect from a second language lies the possibility of a simple transfer effect from English, since all participants were native English speakers. In fact, there exist several English constructions that allow an adjective to follow the noun it modifies in the linear string. These include a special construction involving the word, *thing* (4a and b), resultative and depictive predicates such as those in (5a and b), as well as the postnominal appearance of longer or conjoined adjectives (6a and b):

- (4a) He likes *all things linguistic*    [[*all things*]<sub>NP</sub> Adj]  
 (4b) He owned *something blue*.    [[*something*]<sub>NP</sub> Adj]
- (5a) He saw *the man*                    Depictive secondary  
       *naked*.                                predicate  
 (5b) He hammered *the*                Resultative secondary  
       *metal flat*.                         predicate
- (6a) *The man, big and hairy*, scared    Heavy  
       *the child*.                            postnominal  
 (6b) *The girl, asleep on the couch*,    APs  
       *didn't hear her alarm*.

In all of these cases, the adjectives modify nouns they follow, just as the adjectives modify the nouns in CSL's experiment.

It is difficult to estimate how common these constructions are. Neither the 100 million word BNC nor the 450 million word COCA corpus allows searches of "N Adj" because the number of nouns and adjectives are independently too large. The results of searches for more specific examples of both N Adj and N Num are provided in Table 2.<sup>4</sup> Clearly this provides just a subset of what exists for N Adj, since the resultative construction, for example, occurs with hundreds of verbs, not just with *make* and *find*. As is evident from Table 2, both the types and tokens of N-Adj constructions with the relevant semantics vastly outnumber those of N-Num.

CSL note that numerals may also follow nouns with some regularity (p. 323). In order to examine such cases, I performed a search on the 100 million word tagged BNC corpus for all "N Num" strings. The top hundred types comprised 25,138 tokens, but these types were *all* examples such as *number one*, *chapter 2*, *Table 2*, *War II*, and *point five*. That is, all of these cases designated proper names; none of them quantified the noun. In fact, if a numeral greater than one were to quantify the preceding noun, we would expect the noun to be plural, but not a single

<sup>3</sup> Possible effects of L2 and L1 on a third language change in a complex way. For example, since bilinguals typically find it easier to learn a third language than monolinguals do (e.g. Albert & Opler, 1978; Jacobsen & Imhoof, 1974; Lerea & Kohut, 1961), it is possible that mastery of Spanish could lead to better learning and less transfer.

<sup>4</sup> Approximate values are included for larger numbers of N-A because not all tokens were examined by hand to determine whether they had the appropriate modification semantics. Instead, a random sample was examined and the percentage of examples with the intended semantics was multiplied by the total number of examples. A lower bound estimate is provided.



**Table 2**

Frequencies of a small sample of constructions that allow N-<Adj> order with a modification meaning and an attempt to find *all* of the constructions that allow N-<Num> order with quantification meaning (searches performed on 450 million word COCA corpus on 8/7–17/2012, 12/6–13/2012).

Sample of <N><Adj> constructions	# Of instances with modification interpretation (lower bound given)	Potential <N><Num> constructions	# Of instances with quantification interp.
<i>All things</i> <adj>	≈300 <sup>a</sup>	<i>All things</i> <Num>	0
<i>Something</i> <adj>	≈25,000	<i>Something</i> <Num>	0 <sup>b</sup>
		<i>Some things</i> <Num>	0
<i>Made it</i> <adj>	≈6000	<i>Made them</i> <Num>	0
[find]<pronoun><adj>	≈4000	[find]<pronoun><Num>	0
Heavy postposed construction (sample list)		Heavy postposed construction (possibly exhaustive list)	
<i>N</i> <adj> <i>than</i>	≈2000	<i>N<sub>pl</sub></i> <Num> <i>in all</i>	3
<i>N</i> <adj> <i>with</i>	≈2000	<i>N<sub>pl</sub></i> <Num> <i>at a time</i>	200
<i>N</i> <adj> <i>of</i>	≈4000	<i>N<sub>pl</sub></i> <Num> <i>in number</i>	5

<sup>a</sup> Conservative estimates are provided (see note 4).

<sup>b</sup> There were 138 examples of the linear string, *something* <numeral>, but these *all* involved phrases like, *pulled over for something 22 times*; *built something sixteen feet high*; i.e., none of the numerals quantified *something*. If a postnominal numeral greater than 1 quantifies the noun, the noun should appear in the plural. But there were in fact no tokens of the plural *some things Num*.

one of the nouns in the first 100 types was plural. It is therefore possible that the N-Num/Proper Name construction *interferes* with participants' learning of the N-Num quantificational pattern.

In a review of this article before publication, Jenny Culbertson cites the following examples found on Google, which do not involve proper names:

(7) "Jonah is obsessed with *all things five* since he is five now." [cawfee98.blogspot.com/2008/.../are-you-smarter-than-fifth-grader](http://cawfee98.blogspot.com/2008/.../are-you-smarter-than-fifth-grader).

(8) "... making *the number of episodes 18*." [en.wikipedia.org/wiki/Lost\\_\(season\\_6\)](http://en.wikipedia.org/wiki/Lost_(season_6))

(9) "These dogs, *three in number*..." [www.sacred-texts.com/nec/hpl/hpl27.htm](http://www.sacred-texts.com/nec/hpl/hpl27.htm)

But notice that in (7), Jonah is not obsessed with five things (a quantification interpretation) but with things that come in sets of five. In (8) as well, 18 is the *number* in question; the sentence does not refer to 18 *numbers*; also, one cannot say, \**Episodes eighteen were watched*.

As shown in Fig. 2, there is one English construction that does allow postnominal numerals with a genuine quantification interpretation and it occurs with both adjectives and numerals: longer modifiers/quantifiers can occasionally be postposed. This is exemplified by the "N Num *in number*" expression in (9). But while the postnominal option is generally possible for longer adjectival phrases, there are only a few collocations that allow numerals after nouns. The only ones I could find were N <Num> *in number*, *in total* and *at a time*. As Table 2 indicates, the postnominal numeral construction exemplified by these cases is quite rare compared with postnominal adjectives in the COCA corpus.

One might object that adjectives are simply more frequent than numerals. And indeed, a search of COCA reveals that there almost twice as many tokens of the 100 most frequent adjectives than the 100 most frequent numerals. Assuming a Zipfian distribution whereby the most frequent types account for most of the tokens, we can conclude that adjectives are roughly twice as frequent overall. The same ratio is found in the 100 million word BNC corpus. The 2–1 base rate difference likely plays some

role in the fact that postposed modifying adjectives outnumber postposed quantifying numerals. But the 2:1 difference in base-rates cannot account for the roughly 215:1 skew in favor of postnominal adjectives that is evident in Table 2.

Thus the idea is that the pattern exemplified by *grifta blue* is supported by formally and semantically analogous constructions evident in: *She painted it blue*; *She found it blue*; *I bought something blue*; *the flag, bright red, white and blue*; *she likes all things blue*. In all of these cases the adjective follows and modifies the noun phrase. This offers the opportunity for positive transfer from English to the novel language. On the other hand, *grifta three* is only potentially related to examples like *griftas, three in all*; this type of case is extremely uncommon by comparison, and requires plural marking on the noun. English overwhelmingly prefers numerals following nouns to receive a proper name interpretation rather than a quantificational one (e.g., *Spiderman 3*, *Chapter 3*, *Table 3*, or *Thing II*), and this may have inhibited the intended quantificational interpretation in the experiment (see Ringbom, 1992 for discussion of positive and negative transfer). These transfer effects are only evident, though, clearly, when a simple rule-like generalization (nouns first; or nouns last) cannot be used.

The authors raise the possibility of a transfer effect from English (p. 323), but they reject the idea because it is insufficient to account for *all* the data. In particular, it does not account for the fact that consistent/harmonic grammars are learned more easily than inconsistent grammars. This is clearly true, but what is at issue is whether a domain-specific innate bias against A-N + N-Num order is needed to specifically account for the difference between conditions 3 and 4, or whether instead this difference can be attributed to familiarity with minor constructions of English or with type 3 languages.<sup>5</sup> The

<sup>5</sup> CSL argue that a transfer account cannot explain the preference for condition 3 over condition 4 on the grounds that only temporary ("stage level") properties can appear postnominally in English; they note that the experimental adjectives encoded permanent ("individual level") properties. But examples such as (4a–b), (5b) and (6a) contradict such a claim, and even if it were true, it remains likely that undergrads are able to generalize across temporary and permanent properties, as they are familiar with the category "adjective".

(domain-general) preference for consistency is needed on either account.

That is, CSL demonstrate clearly that participants do not simply prefer English-like data across the board, since they preferred the non-English but consistent order N-A & N-Num (condition 2) to the closer-to-English N-A & Num-N order (condition 3). Thus consistency matters more than familiarity does, at least for adult participants who are likely to use conscious strategies. That is, if learners only need to learn to position the noun first (condition 2), they are quite adept at doing so, even if the pattern differs from their native language. But it is not surprising that the pattern that is neither internally consistent *nor* familiar—the pattern used in condition 4 – is the one that participants found most difficult. The fact that participants in condition 4 found the placement of numerals particularly challenging, not adjectives, further supports the idea that it is English's almost total lack of postnominal numerals as quantifiers that makes condition 4's mini-language particularly strange for English speakers to learn.

## 6. Conclusion

Are the speakers of the 25 languages that actually prefer the “dispreferred” A-N + N-Num order at a disadvantage when learning this aspect of their languages? It is not clear. If the typological tendency is simply the result of shared historical roots and language contact, then the answer could be no.

On the other hand, CSL's paper offers welcome empirical data, and avoids the assumption that virtually *all* facts about syntax are the result of universal, unlearned, non-functional knowledge or biases (e.g., Chomsky, 1981; Bickerton, 1984). With CSL, it seems the reason participants preferred the consistent, harmonic orders over other orders is likely because the consistent orders were simpler. They required the nouns to always occur first (condition 1) or always occur last (condition 2) in the two-word strings.

When the order of the nouns varied between adjectives and numerals, English speakers found the Num-N + N-A order (condition 3) easier than the N-Num + A-N order (condition 4). CSL argue that the A-N + N-Num dispreference is a universal, unlearned domain-specific bias, and the present discussion does not prove that it is not. But an explanation for such a bias would be required and the possibilities surveyed here are not compelling. The present discussion suggests instead that participants were vastly more familiar with N-A order being associated with the relevant modification interpretation than they were with N-Num order having quantificational semantics. Since A-N + N-Num order is neither consistent nor familiar, it was the least likely to be generalized.

## Acknowledgments

I am very grateful to Jenny Culbertson, Bill Croft, Laura Michaelis, and Ivan Sag for very helpful e-mail discussions and to Bill Croft, Sam Glucksberg, Volya Kapatsinski, Phil Johnson-Laird, and Clarice Robenalt for very useful feedback on an earlier draft. In addition, Jenny Culbertson, Luca

Onnis, Matthew Crocker, and a third anonymous reviewer provided insightful reviews of this discussion piece. I'm also grateful to support from the Freie Universität in Berlin for funding through an Einstein Foundation fellowship.

## References

- Abney, S. P. (1987). The English noun phrase in its sentential aspect (Doctoral dissertation, Massachusetts Institute of Technology).
- Albert, M. L., & Obler, L. K. (1978). *The bilingual brain*. New York: Academic Press.
- Bates, E. (1993). *Modularity, domain specificity and the development of language*. Center for Research in Language, UCSD.
- Beckner, C., Blythe, R. A., Bybee, J. L., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system. *Language Learning*, 59(Suppl. 1.1–26).
- Biberauer, T., Holmberg, A., & Roberts, I. (2008). Structure and linearization in disharmonic word orders. In C. B. Chang & H. J. Haynie (Eds.), *Proceedings of the 26th west coast conference on formal linguistics* (pp. 96–104). Somerville, MA: Cascadilla Press.
- Bickerton, Derek. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7, 173–221.
- Blumberg, M. S. (2006). *Basic instinct: The genesis of behavior*. Basic Books.
- 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.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Holland: Foris Publications. Reprint. 7th Edition. Berlin and New York: Mouton de Gruyter, 1993.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The behavioral and brain sciences*, 31(5), 489–508. discussion 509–58.
- Cinque, G. (2005). Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry*, 36, 315–332.
- Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge University Press.
- Croft, William (1996). What's a head? In J. Rooryck & L. Zaring (Eds.), *Phrase structure and the lexicon* (pp. 35–75). Dordrecht: Kluwer Academic Publishers.
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122, 306–329.
- Deák, G. (2000). Hunting the fox of word learning: Why “Constraints” fail to capture it. *Developmental Review*, 20(1), 29–80.
- Dixon, B. M. W. (1977). Where have all the adjectives gone? *Language*, 1(1), 19–80.
- Dryer, M. S. (2000). Word order in Tibeto-Burman languages. *Draft*, 1–62.
- Dryer, M. S. (2008a). Order of adjective and noun. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online* (pp. 354–357). Munich: Max Planck Digital Library.
- Dryer, M. (2008b). Order of numeral and noun. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online* (pp. 362–365).
- Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473, 79–82.
- Ellefsen, M. R., & Christiansen, M. H. (2000). Subagency constraints without universal grammar: Evidence from artificial language learning and connectionist modeling. In *The proceedings of the 22nd annual conference of the cognitive science society* (pp. 645–650).
- Elman, J., Bates, E., et al. (1996). *Rethinking innateness*. MIT Press.
- Falk, Y., & Bardel, C. (2001). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, 27, 59–82.
- Frank, M. C., Everett, D., Fedorenko, E., & Gibson, E. (2008). Number as cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, 108(3), 819–824.
- Gardner, R. A. (1957). Probability-learning with two and three choices. *American Journal of Psychology*, 70, 174–185.
- Goldberg, A. E. (2004). But do we need Universal Grammar? Comment on Lidz et al. (2003). *Cognition*.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: University of Oxford Press.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 58–90). Cambridge, MA: MIT Press.
- Haspelmath, M. (2008). Parametric versus functional explanation of syntactic universals. In T. Biberauer (Ed.), *The limits of syntactic variation* (pp. 75–107). Benjamins.

- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, MA: Cambridge University Press.
- Hudson Kam, C. L., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66.
- Jacobsen, M., & Imhoof, M. (1974). Predicting success in learning a second language. *Modern Language Journal*, 5(8), 329–336.
- Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57(2), 129–191.
- Karmiloff-Smith, A. (1994). Beyond modularity: A developmental perspective on cognitive science. *IJLCD*, 29, 95–105.
- Kayne, Richard S. (1994). In *Antisymmetry*. MIT Press.
- Lakoff, G. (1970). *Irregularity in syntax*. New York: Holt Rinehart & Winston.
- Lerea, L., & Kohut, S. (1961). A comparative study of monolinguals and bilinguals in a verbal task performance. *Journal of Clinical Psychology*, 17, 49–52.
- Polinsky, Maria (2012). In *Headedness again*. Harvard University. [lingbuzz/001583](http://lingbuzz/001583).
- Prince, A., & Smolensky, P. (1993/2004). In *Optimality theory: Constraint interaction in generative grammar*. New York, NY: Blackwell. Revised version published by Blackwell 2004.
- Rangan, K. (1979). *Purki grammar*. Central Institute of Indian Languages.
- Ringbom, H. (1992). On L1 transfer in L2 comprehension and L2 production. *Language Learning*, 42(1), 85–112.
- Simone, R., & Vallauri, E. L. (2010). Natural constraints on language. *CFS*, 63, 205–224.
- Tomasello, M. (2003). *Constructing a Language: a usage-based theory of language acquisition*. Harvard University Press.
- Tomasello, M. (2009). Universal grammar is dead. *BBS*, 32(5), 470–471.
- Weir, M. W. (1972). Probability performance: Reinforcement procedure and number of alternatives. *American Journal of Psychology*, 85, 261–270.
- Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10), 451–456.
- Zwicky, Arnold (1985). Heads. *Journal of Linguistics*, 21, 1–29.