Cognition 204 (2020) 104392

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

A learning bias for word order harmony: Evidence from speakers of non-harmonic languages

Jennifer Culbertson[a,*], Julie Franck[b], Guillaume Braquet[a], Magda Barrera Navarro[a], Inbal Arnon[c]

[a] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, United Kingdom
[b] Department of Psychology, University of Geneva, Switzerland
[c] Department of Psychology, Hebrew University of Jerusalem, Israel

ARTICLE INFO

Keywords:
Word order
Harmony
Syntax
Learning biases
Artificial language learning
Second language learning

ABSTRACT

Word order harmony describes the tendency, found across the world's languages, to consistently order syntactic heads relative to dependents. It is one of the most well-known and well-studied typological universals. Almost since it was first noted by Greenberg (1963), there has been disagreement about what role, if any, the cognitive system plays in driving harmony. Recently, a series of studies using artificial language learning experiments reported that harmonic noun phrase word orders were preferred over non-harmonic orders by English-speaking adults and children (Culbertson et al., 2012; Culbertson & Newport, 2015, 2017). However, this evidence is potentially confounded by the fact that English is itself a harmonic language (Goldberg, 2013). Here we sought to extend the results from these studies by exploring whether learners who have substantial experience with a non-harmonic language still showed a bias for harmonic patterns during learning. We found that monolingual French- and Hebrew-speaking children, whose language has a non-harmonic noun phrase order (N Adj, Num N) nevertheless preferred harmonic patterns when learning an artificial language. We also found evidence for a harmony bias across several populations of adult learners, although this interacted in complex ways with their L2 experience. Our results suggest that transfer from the L1 cannot explain the preference for harmony found in previous studies. Moreover, they provide the strongest evidence yet that a cognitive bias for harmony is a plausible candidate for shaping linguistic typology.

1. Introduction

There are thousands of languages spoken in the world today, and in principle one could imagine that they are all completely distinct from one other. However, it has long been noted that this is not the case. Rather, despite extensive variation, languages exhibit commonalities. The best explanation for these commonalities is one of the most long-standing debates in the study of language, likely in part because they are caused by a multitude of factors. Some languages look similar because they are *genetically* related–they descend from the same mother language (e.g., Italian and Spanish are similar because they both come from Latin; on a much longer timescale, the same is true of English and Sanskrit, which belong to the same language family, Indo-European). Other languages are similar because they are geographically close to one other, and have influenced each other through direct contact (e.g., Heine & Kuteva, 2008; Thomason, 2001). Many researchers also argue that some commonalities arise due to how languages tend to change over time (e.g., Bybee, 2008; Ohala, 1993). Our shared physiology and cognition have also been invoked as possible explanations for similarities between languages. From restrictions on phonological patterns, to the highly skewed distribution of word orders, cross-linguistic patterns across different linguistic domains have been argued to reflect a specialized cognitive mechanism for language (e.g., Cinque, 2005; Harbour, 2016; Hayes et al., 2004). While such explanations have historically been favored by many linguists (e.g., those working in the Chomskian tradition), other researchers in the broader cognitive science community have been skeptical about whether and to what extent such explanations could explain the typological properties of the worlds' languages, after controlling for geographical, and historical factors (e.g., see Evans & Levinson, 2009 and Levinson & Evans, 2010 and accompanying commentaries). Part of this skepticism has been fueled by the lack of empirical evidence: it is very hard to tease apart the effect of the different factors when comparing the languages of the world: at a given point in time, we cannot separate the impact of cognition, history and geography on the structure of a specific language. Over the past decade, however, new experimental paradigms

* Corresponding author at: 3 Charles Street, Edinburgh EH8 9AD, United Kingdom.
E-mail address: jennifer.culbertson@ed.ac.uk (J. Culbertson).

https://doi.org/10.1016/j.cognition.2020.104392
Received 7 November 2019; Received in revised form 14 June 2020; Accepted 26 June 2020
Available online 13 July 2020

have been developed which help move this debate forward. Specifically, researchers have used artificial language learning paradigms, in which adults and children learn miniature constructed languages under highly controlled conditions. While these conditions are unnatural in certain respects (e.g., the languages are highly simplified, and the learning environment differs from how most natural languages are learned), the resulting data can provide much needed corroborating evidence for the role of cognitive biases in shaping typology (for reviews see Culbertson, 2012, Moreton & Pater, 2012a, 2012b). Here we use these experimental methods to evaluate a cognitively-based hypothesis for one of the most well-known cross-linguistic commonalties in syntax: word order harmony. We build on previous, closely-related work, with the goal of showing that both adult and child learners have a bias for harmonic orders, crucially even when their first language does not display this property.

1.1. Word order harmony

Word order harmony describes a well-studied phenomenon in which syntactic heads in a language align with one another across different types of phrases. For example, verbs are heads of verb phrases, which can also include *dependents* like direct objects, e.g., '*kick* [the can]'. Adpositions (like 'up' or 'down' in English) are heads in so-called adpositional phrases, which can also include dependents, e.g., '*down* [the road]'. In many languages, heads across both these types of phrases are ordered either both before their dependents (as in English), or both after their dependents (as in Japanese). The same holds within the noun phrase, where in many languages the noun comes consistently before (as in Thai) *or* after (as in English) different kinds of nominal dependents (or modifiers) like adjectives, numerals, and demonstratives (e.g., '[these] [two] [black] *cats*').[1] Generally, these harmonic patterns tend to be more frequent across the world's languages than non-harmonic ones, where some elements appear before their heads while others appear after.[2] Typological counts of harmonic and non-harmonic combinations of these particular heads and dependents are shown in Table 1.

Beginning with Greenberg's (1963) foundational work on linguistic typology, harmony has a long history in the study of linguistics, and has been approached from a number of theoretical perspectives. Explanations for harmony have included all of the factors described above, from highly specialized, universal constraints on the human linguistic system (e.g., Travis, 1984), to patterns of language change (e.g., Aristar, 1991), to genetic relationships among languages (e.g., Dunn et al., 2011). However, here we focus on two broad classes of explanation which differ in terms of the role they ascribe to the cognitive system. One general class of explanation puts cognition front and center: individual language learners or users prefer harmonic patterns because they are simpler (Vennemann, 1976, Keenan, 1979, Hawkins, 1979, 1983, Mallinson & Blake, 1981, Pater, 2011, Culbertson et al., 2013, and Culbertson & Kirby, 2016), allow easier generalization of word order across phrase types (Baker, 2001; Chomsky, 1988; Travis, 1984), or lead to easier processing (Hahn et al., 2018, Hawkins, 2009). The second class of explanation takes the burden off of individual-level cognition, and instead focuses on common pathways of change: alignment between syntactic heads is due to shared diachrony (Aristar, 1991; Givón, 1975, 1979, 1984; Kaufman, 2009). For example, a common historical source for new adpositions is verbs (Givón, 1975). These two heads will therefore by default share a common order.

1.2. Evidence for a cognitive bias favoring harmony

These two explanations are not necessarily in conflict; both cognitive and historical forces could drive the cross-linguistic over-representation of harmonic orders. However, which is the primary explanation has nevertheless generated extensive discussion among linguists. Indeed, it fits precisely into the major debate outlined above: to what degree do we need to resort to explanations based on cognition to explain why languages look the way they do. Until recently, the debate was largely theoretical: the cognitive explanation was not supported by behavioral evidence showing that individuals in fact prefer harmony either when learning or using language. Culbertson et al. (2012) provided the first such evidence, using an artificial language learning task to investigate whether adult English speakers prefer harmonic order between nouns and different types of nominal modifiers when learning a new language. Below we describe this experiment in some detail, as it provides the basis for the experiments we report here.

Culbertson et al. (2012) taught participants a miniature language in which nouns occurred either with an adjective or a numeral. There was a dominant order for each type of modifier in the language, used most of the time, with the opposite order occurring the remainder of the time. This variation was random noise rather than lexical conditioning (e.g., a particular modifier or noun could occur in any order). Participants were assigned to one of four conditions which differed only in which of the four possible ordering patterns (shown in Table 1B) was the dominant pattern in their input. Previous work in developmental psychology has shown that learners generally do not reproduce random variation like this, but rather reduce the amount of noise, e.g., by picking the most frequent pattern and using it consistently (Ferdinand, Kirby, & Smith, 2019; Hudson Kam & Newport, 2009; Smith & Wonnacott, 2010). This phenomenon is called 'regularization'. In this case, regularization means using the dominant pattern more than it was used in the input. Culbertson et al. (2012) predicted that if learning is predominantly affected by a cognitive bias for harmony (rather than a bias to prefer native-like patterns for example) learners should be more readily able to pick out harmonic dominant patterns from the noise, and therefore should be more likely to regularize them. Importantly, learners were predicted to regularize harmonic patterns regardless of whether both modifiers preceded the noun (as in English), or followed the noun (as in Thai). By contrast, they were predicted not to regularize either non-harmonic pattern. They additionally hypothesized that the non-harmonic pattern Adj N, N Num would be particularly disfavored because it is notably less common than N Adj, Num N cross-linguistically. Culbertson and Newport (2015) replicated this experiment with English-speaking child learners (6–7 years old) to explore how a bias for harmony might change (or not) across development. The experiment was simplified so that children could successfully learn the novel lexicon, but otherwise the same.

In Figs. 1 and 2, we reproduce the results from these two studies. Fig. 1 shows average production of the dominant order across conditions for English-speaking adults and children. Here we have included individual participant data points to illustrate that learners–particularly children–were highly variable in their behavior. In particular, while some matched or regularized the dominant input order, many others produced output distributions that more closely resembled another pattern altogether. Fig. 2 illustrates this by plotting the proportion of pre-nominal order participants produced for each modifier type. If a participant produced a high proportion of pre-nominal orders for both modifier types, then regardless of the input condition, they will be

[1] Note that in the nominal domain, there is some debate concerning whether the noun is the head, taking e. g., adjectives as dependents, or whether instead there are distinct phrases, e.g., AdjP, NumP, DemP, which would have a noun as their dependent (see Culbertson et al., 2012 for additional discussion and relevant references). For our purposes it only matters that across all these types of phrases, whether they contain an adjective, a numeral, or a demonstrative, the ordering of the noun is consistent. If one treats the noun as the head, then harmony in the nominal domain is not a case of harmony across different categories of phrase (as harmony between a verb phrase and an adpositional phrase), but harmony among different types of dependents in a single category, the noun phrase. We return to this in the General Discussion.

[2] Although note that the tendency for harmony across phrases depends on the combination of phrases in question. For example, the order of verb and object does not seem to generally align with the order of noun and adjective (see Dryer 1992). In English the order is non-harmonic: verb-object, but adjective-noun.

Table 1
Counts from a large sample of languages (WALS, Dryer & Haspelmath, 2013) illustrating the dominance of word order harmony between (A) the verb and its dependent object noun and the adposition and its dependent noun, and (B) the noun and an adjectival modifier and the noun and a numeral modifier.

A

	N Verb	Verb N
N Post	472	41
Prep N	14	454

B

	Adj N	N Adj
Num N	251	168
N Num	37	509

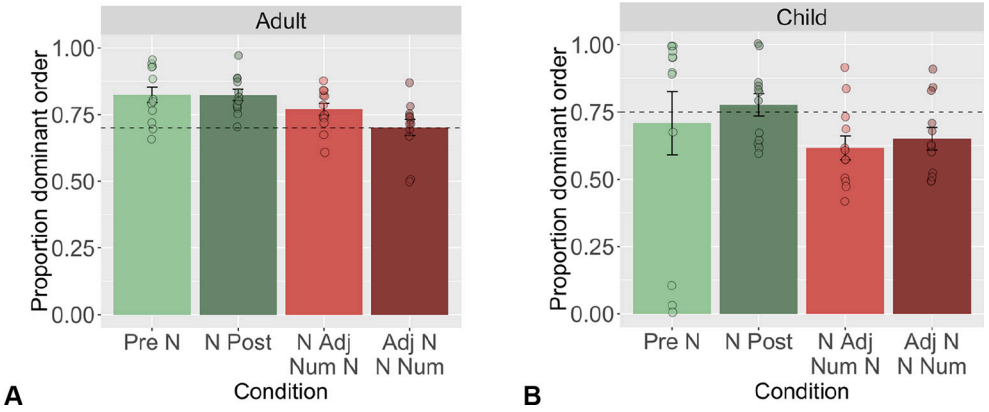

Fig. 1. Proportion use of the dominant order in each condition, for (A) English-speaking adults (Culbertson et al., 2012), and (B) children (Culbertson & Newport, 2015). Bars show group averages, points show individual participants (jittered to prevent overlap), error bars represent standard error on by-participant means. Dotted line shows the frequency of the dominant pattern in the input.

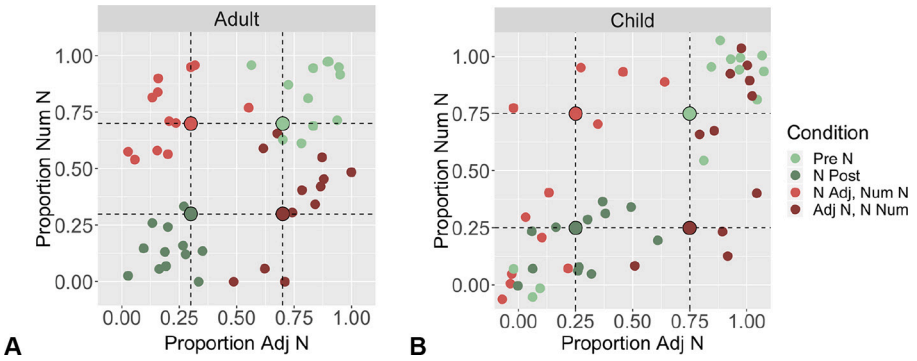

Fig. 2. Individual adult (A) and child (B) participant outcomes. The y-axis indicates proportion use of Num N; the x-axis indicates proportion use of Adj N. The larger filled points at the intersection of the dotted lines represent the input proportions (NB: the dominant order was used 70% of the time for adults, and 75% for children). The corners of this space correspond to deterministic use of one of the four input patterns. The upper right corner corresponds to pre-nominal harmony (Adj N, Num N). The lower left corner corresponds to post-nominal harmony (N Adj, N Num). The upper left corner corresponds to non-harmonic N Adj, N Num. The lower right corner corresponds to non-harmonic Adj N, N Num. Points representing child learners are jittered to prevent overlap.

located in the upper right corner of this space. If they produced a *low* proportion for pre-nominal orders for both modifier types, then they will be located in the opposite corner. Participants whose productions matched one of the two non-harmonic patterns will be in either the upper left or lower right corner.

To summarize these results, English-speaking adult and child learners preferred harmonic to non-harmonic patterns. Adults were more likely to regularize *both* harmonic patterns (no significant difference was found between the two), and no participant regularized the cross-linguistically rare non-harmonic pattern N Adj, Num N. Children's preference for harmony was even stronger: almost all children produced harmonic output patterns, again with no difference between the more English-like Pre N pattern (Adj N, Num N) and the opposite, N Post (N Adj, N Num).[3]

[3] It is worth noting here that the results from Culbertson et al. (2012) were fully replicated in Culbertson and Smolensky (2012). The results from Culbertson and Newport (2015) were also strengthened by a replication with English-speaking children in which the input was deterministically non-harmonic (Culbertson & Newport, 2017). Children *still* produced harmonic output patterns.

1.3. Moving beyond English speakers

Culbertson et al. (2012) and Culbertson and Newport (2015) interpret their results as providing evidence that a cognitive bias favoring harmony drives the cross-linguistic over-representation of harmonic orders. However, Goldberg (2013) counters that the harmony bias English-speaking learners exhibit may reflect transfer. The most straightforward notion of L1 transfer (or influence) would be a clear preference for the English-like Pre N pattern. This was not observed. However, English speakers have experience with a language in which adjectives and numerals behave similarly with respect to their order, therefore they might transfer this expectation to a new language. This kind of abstract transfer would predict both harmonic patterns to have an advantage. Here we investigate how prior linguistic experience might influence learning of nominal word order. The overall aim is to determine whether the harmony bias is still present despite substantial experience with non-harmonic patterns. If the harmonic bias is experience-independent–i.e., detectable even in the face of this kind of experience–then it is a plausible candidate for shaping linguistic typology through learning and use.

In Experiment 1, we test whether monolingual children who speak a *non-harmonic* L1 (either French or Hebrew) exhibit a harmony bias. This is the clearest way to test the predictions of the two alternative accounts of the previous data outlined above. If previous results were due to abstract

transfer of harmony based on experience with a specific harmonic pattern (namely English), then monolingual speakers of a specific non-harmonic language should not show a preference for harmonic patterns. Indeed, they should prefer non-harmonic patterns. By contrast, if previous results were driven by a harmony bias, then we expect to see this same bias regardless of learners' L1. Given the strength of the bias in English-speaking children, combined with the fact that these children are monolingual, Experiment 1 likely gives us the best chance of observing clear results.

In Experiments 2 and 3 we target bilingual adult learners who have substantial experience with both a harmonic language and a non-harmonic one. Part of our motivation for this is practical: the population of adults targeted in previous studies is university students, and in order to match this population while shifting to L1 speakers of a non-harmonic language, we inevitably end up with a bilingual population. However, this gives us the opportunity to explore a more complex question: how adults' prior experience with *multiple language types* might affect their biases in this domain. Research on multilingualism shows that when people learn a new language, they may be influenced not just by their L1, but also by their L2 or any other languages they have substantial fluency in (e.g., Bardel & Falk, 2007, 2012, Rothman et al., 2011, Westergaard et al., 2017). In Experiment 2, we test adults whose L1 is non-harmonic (either French or Hebrew), but who are bilingual in English. In Experiment 3, we flip this around by testing adults whose L1 is English, but who are bilingual speakers of a non-harmonic language (either French or Spanish). If L1 influence–specifically abstract transfer–is the main driver of learners' behavior, then we predict a bias for non-harmonic orders in Experiment 2, but a bias for harmony in Experiment 3. By contrast, if the L2 exerts substantial pressure, then we might expect the opposite pattern of results. However, an experience-independent bias for harmony predicts that we should see evidence of this bias in all populations, regardless of their backgrounds.[4]

2. Experiment 1: child learners with a non-harmonic L1

In Experiment 1, we investigate whether children whose native language uses a non-harmonic order in the noun phrase nevertheless show a harmony bias when learning a new language. We focus on two languages in particular: French and Hebrew. The French data were first reported in Braquet and Culbertson (2017).

Example noun phrases in these two languages are shown in (1) and (2) below. The examples in (1) illustrate the default order for adjectives, i.e., N Adj (a) and the fixed order for numerals, i.e., Num N (c) in French. While most adjectives typically follow the noun, a small lexically-determined set obligatorily precede the noun, as in (b). In addition, many other adjectives may optionally precede the noun, in which case they have an emphatic reading (see Fox & Thuilier, 2012 for a more complete discussion of adjective flexibility in French). The examples in (2) illustrate the fixed order for adjectives, i.e., N Adj (a) and numerals, i.e., Num N (b) in Hebrew. Adjectives obligatorily follow the noun (note that the writing system is right-to-left, therefore in (2a) for example, the word for 'chair', כסא, is read first, then the word for 'red', אדום). All numerals except the number 'one' (derived from an article) precede the noun.

(1) a. chaise rouge
chair red
'red chair'
b. petite chaise
small chair
'small chair'
c. deux chaises
two chairs
'two chairs'

(2) a. אדום כסא
red chair
'red chair'
b. כיסאות שני
chairs two
'two chairs'

(NB: right-to-left writing system)

If previous results reflect abstract transfer of L1 harmonic order to a new language, then these learners are predicted to prefer *non-harmonic* patterns. This may present as a bias for regularizing non-harmonic dominant input orders more than harmonic input orders. However, as noted above child learners in Culbertson and Newport (2015) often produced output patterns that did not match their input, and yet tended to be harmonic (see also Culbertson & Newport, 2017). Therefore a preference for non-harmonic orders in the child populations tested here may be reflected in a tendency to prefer non-harmonic outputs (and to use *them* consistently) regardless of the input. By contrast, if previous results have reflected a universal harmony bias, then we predict a preference for harmonic patterns instead.

2.1. Method

The design of the experiment was modelled closely after Culbertson and Newport (2015). There, children participated in two sessions, during which they were trained and tested on an artificial language. Here, we use the single-session version of this procedure developed in Culbertson and Newport (2017). Participants were randomly assigned to one of four conditions, which differed only in the frequency with which pre- and post-nominal adjectives and numerals were used. Phrases in the language were comprised of *either* a noun and an adjective *or* a noun and a numeral. Each condition had a dominant order for each modifier type, which was used 75% of the time. In two conditions, the dominant order was harmonic, in the remaining two, it was non-harmonic. Variation in order within a given condition was random; it was not conditioned on the particular lexical items in a phrase. Table 2 describes the four conditions. Each child participated in a single 25–30 minute session which included exposure to the language, followed by a critical test in which learners were asked to produce phrases in the language.

Table 2
The four experimental conditions: two harmonic and two non-harmonic. Each condition featured a dominant order, used in 75% phrases for a given modifier type (adjective or numeral) and 25% of the opposite order.

Harmonic		Non-harmonic	
Condition name	Description	Condition name	Description
Pre N	75% Adj N 75% Num N	N Adj, Num N	75% N Adj 75% Num N
N Post	75% N Adj 75% N Num	Adj N, N Num	75% Adj N 75% N Num

Table 3
IPA transcriptions (and meanings for adjectives and numerals) of French and Hebrew artificial language lexicon. Note that adjectives and numerals are pseudo-nonce (real word equivalents and IPA transcriptions in the respective languages are given in parentheses).

French

Nouns	Adjectives		Numerals	
[bogi]	[bly]	'blue' (cf. *blu* [blø])	[doks]	'two' (cf. *deux* [dø])
[sefi]	[taʃu]	'spotted' (cf. *tacheté* [taʃte])	[tʁa]	'three' (cf. *trois* [tʁwa])
[voli]	[pølu]	'furry' (cf. *poilu* [pwaly])	[kitʁ]	'four' (cf. *quatre* [kætʁ])
[kani]				

Hebrew

Nouns	Adjectives		Numerals	
[bugi]	[dadom]	'red' (cf. אדום [adom])	[ʃta]	'two' (cf. שניים [ʃtaim])
[kani]	[dol]	'big' (cf. גדול [gadol])	[loʃ]	'three' (cf. שלוש [ʃaloʃ])
[safi]	[tin]	'small' (cf. קטן [katain])	[arb]	'four' (cf. ארבע [arba])
[neʃu]				

[4] Data from all experiments is provided at https://osf.io/fdkh3/.

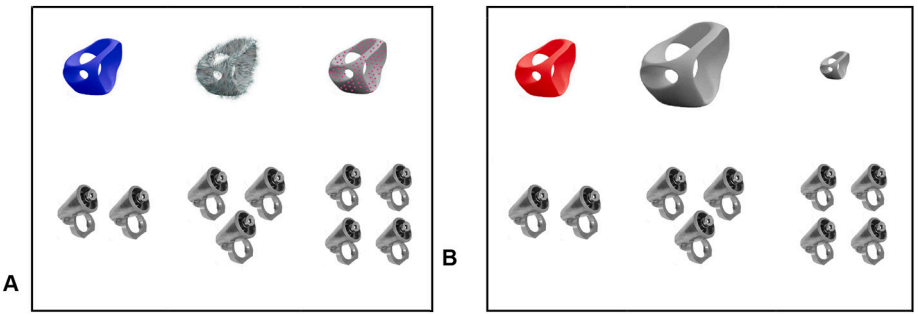

Fig. 3. Example visual stimuli: objects modified by properties (top) or numerosities (bottom). (A) Set used with French-speaking children. (B) Set used with Hebrew-speaking children.

2.1.1. Participants

French-speaking participants were 47 children (24 females), 6–7 years of age (mean = 6;7), recruited from elementary schools in Southwest France. All were native speakers of French, who were either monolingual, or bilingual in French and Occitan (a Romance language spoken in this region, which uses the same nominal word order as French). Hebrew-speaking participants were 43 children (21 females), 6–8 years of age (mean = 7;4), recruited from elementary schools in Israel. All were monolingual. Parental consent was obtained for all participants. Three additional French-speaking children were excluded from the analysis due to failure to complete the full session (2), or extremely low accuracy on the initial noun vocabulary (< 50%) (1). Seven additional Hebrew-speaking children were excluded from the analysis due to failure to complete the full session (3), extremely low accuracy on the initial noun vocabulary (1), technical problems during the experiment (2), or an ADHD diagnosis (1).

2.1.2. Stimuli

Children were taught a language with 4 nouns and 6 modifiers (3 adjectives and 3 numerals). Following Culbertson and Newport (2015, 2017), the lexical items for nouns were fully nonce (representing unfamiliar objects), and the modifiers were pseudo-nonce. Pseudo-nonce modifiers resembled the corresponding Hebrew or French words (i.e., differed by one or two segments). The set of adjectives was different for the two languages. The French set corresponded to the adjectives used in Culbertson and Newport (2015, 2017). The Hebrew set was altered because 'spotted' and 'furry' are longer (in syllables) than the other words. Instead, 'big' and 'small' were used (these were also used in Culbertson et al., 2012). The color word 'red' was used instead of 'blue' because it was simpler to generate a pseudo-nonce form for that term. All lexical items are shown in Table 3.

The visual stimuli were a set of four unfamiliar objects, modified with the properties specified above (adjectives) or grouped according to the numerosities specified above (numerals). Example stimuli are shown in Fig. 3.

2.1.3. Procedure

The experiment was conducted on a laptop computer in a quiet room. The experimenter sat next to the child throughout. Participants were told they would be learning part of a new language with the help of an 'alien informant'. The experimental session progressed in three phases: noun training and testing; phrase training and comprehension; phrase production (critical test).

2.1.3.1. Noun training and testing. Participants were first trained and tested on the noun vocabulary. In the noun training phase, the image of a noun appeared on the screen and the alien provided the label aloud. Participants were instructed to repeat the label aloud. Each noun was repeated 5 times (20 trials total, randomized). Participants then played a noun matching game in which they saw each of the four objects on the screen, heard the alien provide the label for one of them, and were instructed to click on the matching image. A correct response generated a correct feedback sound, and 10 points. An incorrect response generated an incorrect feedback sound. The correct image remained on the screen for 500 ms. Each noun was repeated 5 times (20 trials total, randomized). In the noun testing phase, participants saw an image and were instructed to provide its label aloud. Once they had given their response, the alien provided the correct label (not contingent on the participant's answer). Each noun was repeated 5 times (20 trials total, randomized).

2.1.3.2. Phrase training and comprehension. Participants were then trained on phrases in the language. An image appeared on the screen and the alien provided a phrase to describe it aloud. The image was either an object modified by a property (corresponding to one of the adjectives), or several of the same objects (2, 3, or 4, corresponding to one of the numerals). Participants were instructed to repeat the phrase aloud. Each noun occurred 6 times (twice with each adjective and each numeral, in randomized order, 48 total trials). They were then tested on their comprehension. In comprehension trials, participants saw four images on the screen, heard the alien provide a description, and were instructed to click on the matching image. A correct response generated a correct feedback sound, and 10 points. An incorrect response generated an incorrect feedback sound. The correct image remained on the screen for 500 ms. Each noun occurred 6 times (twice with each adjective and each numeral, in randomized order, 48 total trials).

2.1.3.3. Phrase production (critical test). In production trials, an image appeared on the screen and participants were instructed to describe it aloud. The image was either an object modified by a property (adjective), or several of the same objects (numeral). No feedback was provided. Each noun occurred 6 times (twice with each adjective and each numeral, in randomized order, 48 total trials).

2.2. Results

2.2.1. Coding and vocabulary accuracy

Participants' productions in the critical test phase were coded for order (pre- or post-nominal) by native speakers of French or Hebrew. Vocabulary accuracy scores were also coded for the Hebrew data.[5]

[5] The French data were coded only for order, and original audio recordings were subsequently lost due to malfunction of a backup drive. Recall however, that participants who struggled to learn the noun lexicon were excluded from the analysis. Note that here and throughout an individual lexical item was scored as correct so long as it did not differ from the target by the substitution, addition or deletion of more than one segment. A phrase produced was scored as correct only if both words in the phrase (a noun along with either an adjective or a numeral) were correct. Note that in almost all cases, children produced two word utterances (< 1% involved a missing word), therefore incorrect phrases involved either the noun or modifier (or both) produced incorrectly. This is likely in part because the experimenter would encourage

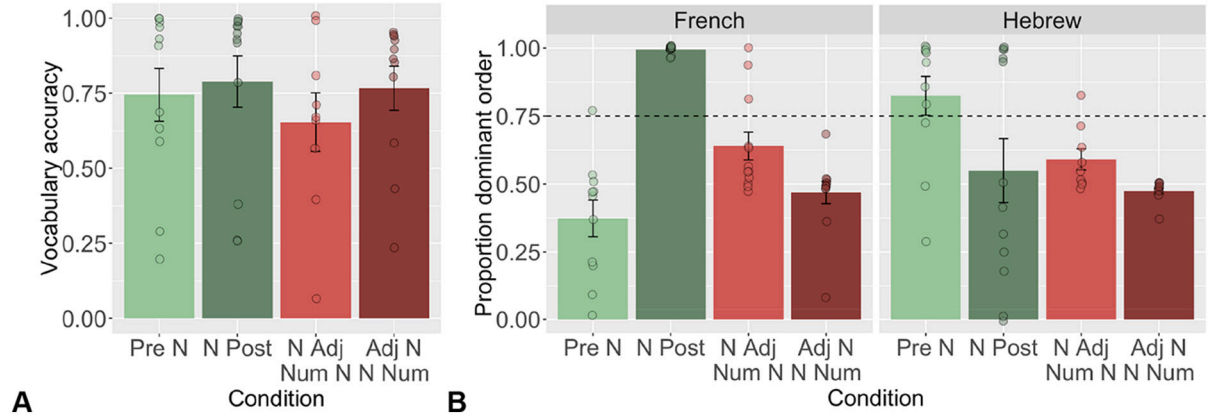

Fig. 4. A: Vocabulary accuracy during critical testing phase for Hebrew-speaking children. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means. B: Proportion use of the dominant order in each condition across French- and Hebrew-speaking children. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means. Dotted line shows the frequency of the dominant pattern in the input.

Fig. 4A shows vocabulary accuracy for this group. To determine whether there were significant differences across input conditions in vocabulary accuracy, we fit two logistic mixed-effects regression models.[6] The first was a null model with only an intercept term. The second was a model including condition as a predictor. We then compared these models using likelihood ratio tests.[7] Adding condition as a predictor did not significantly improve the fit of the model (χ^2 = 1.96, p = 0.58). This indicates that any differences in the production of *order* across input conditions cannot be explained based on level of fluency with the novel lexicon.

2.2.2. Use and regularization of the dominant input order

Fig. 4B shows average production of the dominant order across conditions for each group. As this figure makes clear, participants did not consistently regularize the dominant input pattern. For French-speaking children, only N Post was used in significantly more than 75% of productions (t(11) = 65.23, p < 0.001, all other conditions were either marginally or significantly *below* 75%). For Hebrew-speaking children, the dominant order was not used in significantly more than 75% of productions in any of the conditions (for Pre N, t(10) = 1.05, p = 0.32, all other conditions were either marginally or significantly *below* 75%). We will return to the issue of children's regularization below.

To determine whether there were significant differences in the use of the dominant input order *across* conditions, we fit two logistic mixed-effects regression models for each group. The first was a null model with only an intercept term. The second was a model including condition as a predictor. We then compared these models using likelihood ratio tests. For both groups of participants, adding condition as a predictor significantly improved the fit of the model (French: χ^2 = 80.56, p < 0.001; Hebrew: χ^2 = 10.80, p = 0.01). To probe differences between conditions in each group further, all conditions were compared to each other.[8] In the French-speaking group, N Post differed significantly from all other conditions, with participants in that condition being more likely to use their dominant input order (vs. Pre N: β = 6.3 ± 0.73, p < 0.001; vs. N Adj, Num N: β = 4.9 ± 0.73, p < 0.001; vs. Adj N, N Num: β = 6.3 ± 0.73, p < 0.001). In addition, participants in the Pre N condition were significantly less likely to use their dominant order than participants in N Adj, Num N (β = −1.40 ± 0.41, p = 0.004) but not Adj N, N Num (β = −0.5 ± 0.41, p = 0.59). Finally, use of input order in N Adj, Num N and Adj N, N Num conditions did not differ significantly (β = 0.89 ± 0.40, p = 0.11). In the Hebrew-speaking group, the only significant difference was between Pre N and Adj N, N Num (β = 2.74 ± 0.84, p = 0.006), with participants in Pre N being more likely to use the input order. The difference between Pre N and the other two conditions was marginal (v. N Adj, Num N: β = 2.23 ± 0.88, p = 0.05; v. N Post: β = 1.98 ± 0.83, p = 0.08). No other differences approached significance (highest β = 0.76 ± 0.81, lowest p = 0.78).

2.2.3. Preferred patterns

While analyses of children's production of the input order gives us some indication of which input pattern learners were more likely to match–N Post for the French children, and Pre N for the Hebrew children–it does not in fact tell us much about which patterns learners preferred to use, independently of the input they were exposed to. It also gives the false impression that learners were not generally systematic in their productions (see next section). This is because, as in previous studies, children did not always use the pattern they were trained on. Instead, many learners dramatically shifted their output productions in a way that more closely resembled another pattern altogether. Fig. 5A therefore provides a better illustration of the distribution of patterns participants produced. This figure makes it clear that learners across both populations tended to shift toward one of the two harmonic patterns (specifically N Post in the French group), regardless of their input. Not only do they not regularize the (non-L1-like) non-harmonic pattern Adj N, N Num when it is the dominant pattern in

(*footnote continued*)
children to produce two words–carefully, without introducing any clues about order (e.g., by saying 'anything else'?)–if they did not spontaneously.

[6] All regression models reported here were run using the lme4 package in R (Bates, 2010). All models include random by-participant and by-noun random intercepts where possible. In many cases, adding the latter resulted in singular model fit warnings, therefore we chose to report the models without them. However, in no case did the models including by-noun random intercepts result in different patterns of significance. Where by-noun random intercepts were included, we also attempted to fit models including by-condition random slopes for items, however no such models successfully converged. For each regression analysis reported here, model output tables showing additional details are provided at https://osf.io/fdkh3/.

[7] Likelihood ratio tests are a standard method for nested model comparison, i.e., where one model is a subset of the other and we are interested in testing whether the added complexity of the larger model is warranted. These were conducted using the anova() function in the base stats package in R (R Development Core Team, 2010).

[8] Pairwise comparisons were computed using the multcomp package (Hothorn et al., 2008), using Tukey's method of correcting for multiple comparisons.

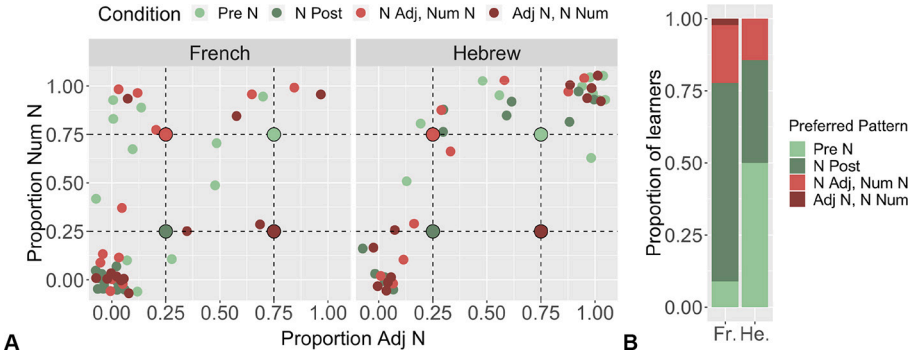

Fig. 5. A: Individual participant outcomes for French- and Hebrew-speaking children distributed across the space of possible outcomes. Here the y-axis indicates the proportion use of Num N, and the x-axis indicates the proportion use of Adj N. The corners of this space correspond to perfectly deterministic use of one of the four input patterns. The upper right corner corresponds to pre-nominal harmony (Adj N, Num N). The lower left corner corresponds to post-nominal harmony (N Adj, N Num). The upper left corner corresponds to non-harmonic (N Adj, N Num). The lower right corner corresponds to non-harmonic (Adj N, N Num). Note that points are jittered to prevent overlap. B: Proportion of learners in each language group (French and Hebrew) whose preferred pattern matches each of the four possible patterns.

their input, but they do not shift to this pattern either. Indeed, classifying the preferred pattern of each child based on which order (pre-nominal or post-nominal) they used most for each modifier type yielded an overwhelming majority of harmonic choice (71/90 of which 46 were N Post), compared to non-harmonic (16/90, of which 15 were L1-like N Adj, Num N), as confirmed by a two-tailed binomial test ($p < 0.001$).[9] This is summarized in Fig. 5B.

2.2.4. Consistency of use of preferred patterns

Once we have determined which patterns individual learners prefer, we can also ask to what degree do they consistently use those patterns. For example, a learner who was trained on N Adj, Num N but prefers N Post might use that pattern in a way that is perfectly consistent (i.e., in nearly all productions), or they may produce that pattern only noisily (e.g., in 60% of productions for each modifier type). This is quantified in Fig. 6. Interestingly, the same patterns which learners were more likely to prefer in each group (i.e., N Post in French, both harmonic patterns in Hebrew) were also used with a greater degree of consistency. We analyzed consistency of use depending on the preferred pattern using mixed-effects logistic regression models and model comparison. For both groups of participants, adding preferred pattern as a predictor significantly improved the fit of the model (French: $\chi^2 = 15.98$, $p = 0.001$; Hebrew: $\chi^2 = 12.51$, $p = 0.002$). To probe differences between conditions in each group further, all conditions were compared to each other (Adj N, N Num was removed, since this was the preferred pattern for only a single French participant). In the French group, N Post differed significantly from N Adj, Num N ($\beta = 2.23 \pm 0.71$, $p = 0.008$), but only marginally from Pre N ($\beta = 2.24 \pm 0.94$, $p = 0.07$); Pre N did not differ from N Adj, Num N ($\beta = -0.02 \pm 0.99$, $p = 1.00$). In the Hebrew group, both N Post and Pre N differed significantly from N Adj, Num N (N Post vs.: $\beta = 2.42 \pm 0.67$, $p < 0.001$; vs. Pre N vs.: $\beta = 1.81 \pm 0.62$, $p = 0.008$), and N Post and Pre N did not differ ($\beta = 0.60 \pm 0.51$, $p = 0.47$).

2.3. Discussion

In this experiment, we taught child learners of two non-harmonic L1s–French and Hebrew–an artificial language with one of four dominant word order patterns. Two patterns were dominant harmonic, and two were dominant non-harmonic. We predicted that if there is a general cognitive bias for harmony, these learners, like English-speaking children, would prefer harmonic to non-harmonic patterns. This could manifest as a preference to regularize harmonic input patterns more than non-harmonic patterns, or as a more general preference to use a harmonic pattern regardless of the input (as in Culbertson & Newport, 2015, 2017). By contrast, if previous evidence of a harmony bias in English children is due to abstract transfer, this would predict a general preference for *non-harmonic* patterns. The latter prediction was clearly not borne out. Rather, our results are broadly consistent with a harmony bias across these two populations. Regardless of the input patterns they were trained on, learners overwhelmingly produced harmonic outputs, and did so with a high level of consistency. When the pattern they produced most was non-harmonic, it was (almost) always the one that matched their native language–N Adj, Num N (though see General Discussion for an alternative interpretation).

Interestingly, the two language populations did differ in the degree to which they used the two harmonic patterns. Hebrew-speaking learners were more likely to regularize the input pattern they were taught if it was Pre N. However, analysis of their preferred patterns indicated a roughly even distribution between Pre N and N Post. In addition, when they preferred either harmonic pattern they were highly consistent in using both. By contrast, French-speaking learners favored N Post: they regularized it when it was the dominant order they were trained on, and were also more likely to use it, and use it consistently, as their preferred pattern. One obvious difference between these two populations is the flexibility in adjective ordering found in French. Intuitively this might lead to the expectation that French-learners would be *more* likely to allow for the possibility of pre-nominal adjectives (and thus perhaps pre-nominal harmony), but this is not what appears to happen. Perhaps French children have a heightened awareness that certain adjectives are constrained to appear post-nominally.[10] For example, color and texture terms such as the ones used in this study invariably appear post-nominally in French. By contrast, pre-nominal order is actually fairly frequent in French, since the subset of adjectives that only appear pre-

[9] Recall that participants produce a single modifier, adjective or numeral, in each utterance. To classify the pattern each participant used the most, we calculated, for each modifier type, whether they were more likely to use pre- or post-nominal order (> 50% of the time). We then combined these two calculations to determine each participants' preferred order. For example, if a child used N-Adj in 65% of utterances, and Num-N in 73% of utterances, they would be classified as preferring N-Adj, Num-N. For three children, a classification could not be determined by this metric (i.e., productions featured exactly 50% of each order for one or both modifier types).

[10] French-speaking children acquire the order of adjectives and numerals early in their language. We extracted all instances of noun phrases including an adjective or one of the numeral words 'two' through 'ten' ('one' in French, *un (e)*, corresponds to the indefinite article, and therefore may be acquired differently from other numerals) from the Lyon corpus (Demuth & Tremblay, 2008). This is a publicly available corpus of naturalistic interactions of 5 parent-child dyads, recorded for 1 h every 1–2 weeks from age 1 to 3 years (185 h of speech total). Children's first noun phrases with both a numeral and an adjective occurred as early as 1;9. Out of a total of 258 instances of numerals modifying a noun, no word order errors were found. Out of 704 instances of adjectives modifying a noun, 6 errors were found (all from a single child, involving post-nominal placement of obligatorily pre-nominal adjectives). See Braquet and Culbertson (2017) for additional discussion. We can therefore say that at age 6–7 French children have long-since mastered the basic order of nominal modifiers in their language. Of course, this is not direct evidence of heightened awareness about adjective position relative to e.g., Hebrew-speaking children.

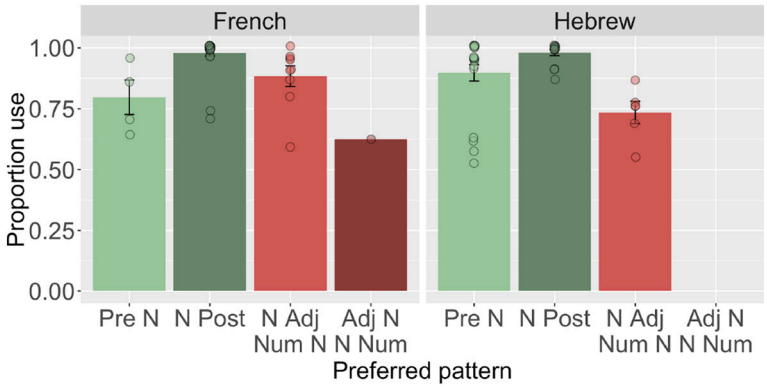

Fig. 6. Proportion use of each preferred pattern for French- and Hebrew-speaking children. For example, the Pre N bar groups all participants whose most frequently used pattern is Pre N, and shows the proportion of time they used that pattern. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means. Note that only one (French) participant used Adj N, N Num as their chosen pattern.

nominally are highly frequent adjectives like 'good', 'big', 'new'.[11] Children may therefore have to actively inhibit pre-nominal order usage for those adjectives which can only occur post-nominally. If French speaking children's awareness of lexically-based constraints on adjective order influences leads to stronger expectations about placement of these adjectives, then among the two harmonic patterns, Post N fulfills this.

To summarize, the results from Experiment 1 provide clear evidence that harmony of the L1 is not what drives the harmony bias found in English-speaking children. In Experiment 2 we aim to provide such evidence for adult speakers of French and Hebrew.

3. Experiment 2: adult learners with a non-harmonic L1

In Experiment 2, we tested adult native speakers of French and Hebrew. In both cases, populations matched to the native English-speaking adult learners tested in Culbertson et al. (2012)–that is, university students–are almost uniformly bilingual in English (and in some cases, other languages as well). This means that our participants will have substantial experience with both harmonic and non-harmonic noun phrase word orders. Nevertheless, testing these populations will shed light on whether the harmony bias changes in native speakers of a non-harmonic L1, as their linguistic experience changes. Based on Culbertson and Newport (2015), we can predict that adults will generally be more successful at matching the input than children (e.g., compare Fig. 2A and B). Therefore, they may be more likely to produce non-harmonic patterns when exposed to them. On top of this, adults' substantial experience with a non-harmonic L1 could lead to a preference for such patterns over non-harmonic ones (e.g., as suggested by Goldberg, 2013), in contrast to children's preferences in Experiment 1. However, evidence from research on bilingualism also suggests that learners' L2 may play a role in their learning of a new language (e.g., Bardel & Falk, 2007, 2012, Rothman et al., 2011, Westergaard et al., 2017). If the L2 substantially influences learning of novel languages, then the pattern of results may be more complex. Specifically, since the dominant L2 for all our adult participants was English, we may expect them to show a preference for English or English-like (i.e., harmonic) order.

3.1. Method

The design of the study was modelled closely after Culbertson et al. (2012), it is also very similar to that described above for Experiment 1. Participants were randomly assigned to one of four conditions, which differed only in the frequency with which pre- and post-nominal adjectives and numerals were used. Phrases in the language were comprised of *either* a noun and an adjective *or* a noun and a numeral. Each condition had a dominant order for each modifier type, which was used 70% of the time. In two conditions, the dominant order was harmonic, in the remaining two, it was non-harmonic. Variation in order within a given condition was random; it was not conditioned on the particular lexical items in a phrase. Each adult participated in a single 45–60 minute session which included exposure to the language, followed by a critical test in which learners were asked to produce phrases in the language.

There was one major difference from Culbertson et al. (2012). In the original study, participants were given immediate online feedback contingent on their productions, including during the 80 critical trials in which they produced phrases in the new language. Culbertson and Smolensky (2012), report a replication in which no such feedback was given, and an additional change to the method was made so that participants were not required to produce 80 phrases without any chance to recall vocabulary items they had trouble with. Participants instead completed several rounds which alternated between comprehension and production. This did not change the outcome of the experiment in Culbertson and Smolensky (2012), therefore we follow this method here, as described in detail below.

3.1.1. Participants

Native French-speaking participants were 88 adults, recruited from the student population at the University of Geneva. Native Hebrew-speaking participants were 74 adults, recruited from the student population at the Hebrew University. All participants completed a version of the LEAP-Q language background questionnaire (Marian et al., 2007). We used self-reported proficiency in reading, speaking, and understanding to create a composite average proficiency score. A histogram showing the distribution of English proficiency scores among French- and Hebrew-speaking participants is shown in Fig. 7.

3.1.2. Stimuli

Adults were taught a language with 10 nouns and 10 modifiers (5 adjectives and 5 numerals). All lexical items were fully nonce. Nouns were two- or three-syllable words that consistently ended in 'i'.[12] Modifiers were single-syllable words. All lexical items are shown in Table 4.

The visual stimuli were a set of ten unfamiliar objects, modified with the properties specified above (adjectives) or grouped according to the numerosities specified above (numerals). Stimuli were a superset of those used for Experiment 1 (see Fig. 3), and are identical to those used

[11] These adjectives are particularly common in speech to children. For example, in the Lyon corpus, 93% of adjectives used by mothers to children are pre-nominal. Although it is notable that of the remaining post-nominal cases, there are relatively many types. By contrast, in the French adult-direct speech corpora available in the Universal Dependencies Treebank, only 28% of adjectives are pre-nominal (though these are most written corpora, so the divergence is likely exaggerated; McDonald et al., 2013).

[12] For reference, Table 5 below shows the English lexicon.

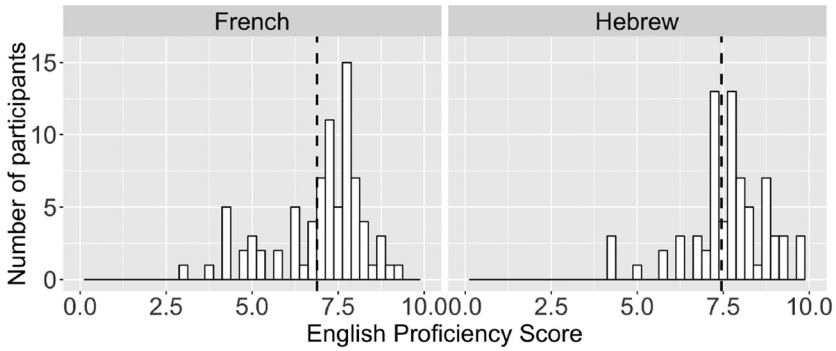

Fig. 7. Distribution of English proficiency composite scores for French- and Hebrew-speaking adults. Dashed line shows mean for each group.

Table 4
IPA transcriptions (and meanings for adjectives and numerals) of the artificial language lexicon used for French and Hebrew participants respectively.

French					
Nouns		Adjectives		Numerals	
[gʁasti]	[klaməgi]	[pʁal]	'big'	[stɔ̃k]	'two'
[vyʁti]	[ʁuspani]	[blun]	'small'	[fʁiʒ]	'three'
[flaʁbi]	[dapəgi]	[guz]	'green'	[nɛ̃p]	'four'
[mutʁi]	[təfɔ̃di]	[dɛ̃s]	'blue'	[zam]	'five'
[brɛfozi]	[pɔnaʁli]	[ʃɑ̃gʁ]	'furry'	[ʒɔl]	'six'

Hebrew					
Nouns		Adjectives		Numerals	
[grabli]	[slebugi]	[dol]	'big'	[suk]	'two'
[vansi]	[rislumi]	[tin]	'small'	[fib]	'three'
[flungi]	[dapegi]	[git]	'green'	[neb]	'four'
[menʃi]	[tapuni]	[dis]	'blue'	[zim]	'five'
[bafuni]	[panarsi]	[ʃaz]	'furry'	[val]	'six'

in Culbertson et al. (2012).

3.1.3. Procedure

The experiment was conducted on a desktop or laptop computer in a quiet room. Participants were told they would be learning part of a new language with the help of an 'alien informant'. The experimental session progressed in three phases: noun training and testing; phrase training; phrase comprehension and production (critical test).

3.1.3.1. Noun training and testing. Participants were first trained and tested on the noun vocabulary. In the noun training phase, the image of a noun appeared on the screen and the alien provided the label aloud. The label was also presented orthographically above the image. Participants were instructed to repeat the label aloud. Each noun was repeated 5 times (50 trials total, randomized). In the noun testing phase, participants saw an image and were instructed to provide its label aloud. Once they had given their response, the alien provided the correct label (not contingent on the participant's answer). Each noun was repeated 5 times (50 trials total, randomized). All participants went through a second round of noun training and testing to ensure they had learned the labels.

3.1.3.2. Phrase training. Participants were then trained on phrases in the language. In the phrase training phase, an image appeared on the screen and the alien provided a phrase to describe it aloud. The image was either an object modified by a property (corresponding to big, small, green, blue, fuzzy), or several of the same objects (corresponding to two, three, four, five, or six). The phrase was also presented orthographically above the image. Participants were instructed to repeat the phrase aloud. Each noun occurred 8 times (with 4 different adjectives and 4 different numerals, in randomized order, 80 total trials).

3.1.3.3. Phrase comprehension and production (critical test). Finally, participants were tested on their comprehension and production in 8 alternating blocks of 20 trials each (4 comprehension, 4 production, 80 total trials of each type). In comprehension trials, participants saw four images on the screen, heard the alien provide a description, and were instructed to click on the matching image. Descriptions were also presented orthographically above the image set. A correct response generated a correct feedback sound, and 10 points. An incorrect response generated an incorrect feedback sound. The correct image remained on the screen for 500 ms. Each noun occurred 8 times (with 4 different adjectives and 4 different numerals, in randomized order). In production trials, an image appeared on the screen and participants were instructed to describe it aloud. The image was either an object modified by a property, or several of the same objects. No feedback was provided. Each noun occurred 8 times (with 4 different adjectives and 4 different numerals, in randomized order).

3.2. Results

3.2.1. Coding and vocabulary accuracy

Participants' productions in the critical test phase were coded for vocabulary accuracy and order (pre- or post-nominal) by native speakers of French or Hebrew.[13] Fig. 8 shows vocabulary accuracy across conditions for each group. To determine whether there were significant differences across conditions in vocabulary accuracy, we fit two logistic mixed-effects regression models for each group. The first was a null model with only an intercept term. The second was a model including condition as a predictor. We then compared these models using likelihood ratio tests. Adding condition as a predictor did not significantly improve the fit of the model for either group of participants (French: $\chi^2 = 2.27$, $p = 0.52$; Hebrew: $\chi^2 = 4.07$, $p = 0.25$).

3.2.2. Use and regularization of the dominant input order

Fig. 9 shows average production of the dominant order across conditions for each group. As this figure suggests, for both groups, only participants whose dominant input order was Pre N regularized above the input level (French: $t(20) = 3.65$, $p = 0.001$; Hebrew: $t(16) = 2.40$, $p = 0.03$). To determine whether there were significant differences across conditions in use of the dominant input order, we fit

[13] Following Culbertson et al. (2012) we discarded trials in which the vocabulary was incorrect. This resulted in the exclusion of approximately 15% of trials. In almost all cases, participants produced two word utterances (< 5% involved a missing word), therefore incorrect phrases involved either the noun or modifier (or both) produced incorrectly. However, note that including these trials does not have any substantial effect on the results reported here.

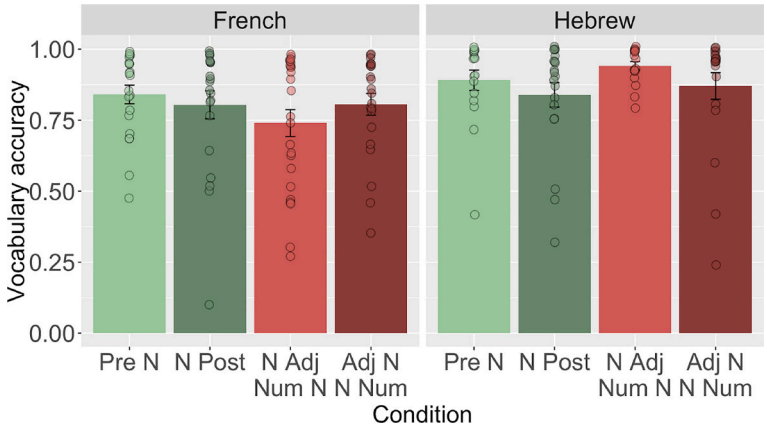

Fig. 8. Vocabulary accuracy during critical testing phase across conditions for French- and Hebrew-speaking adults. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means.

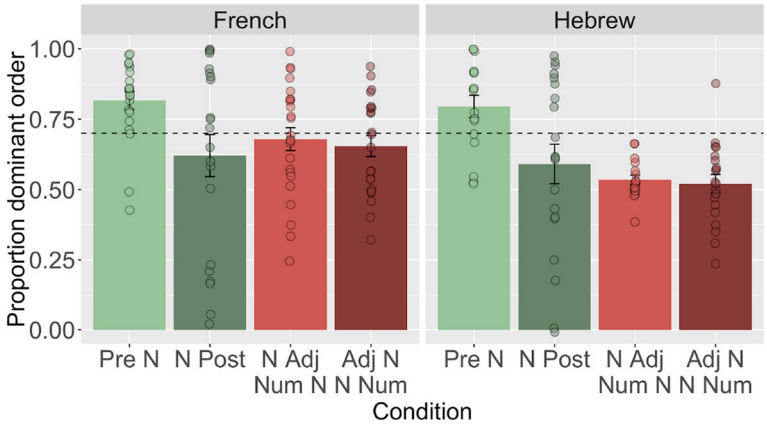

Fig. 9. Proportion use of the dominant order in each condition for French- and Hebrew-speaking adults. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means. Dotted line shows the frequency of the dominant pattern in the input.

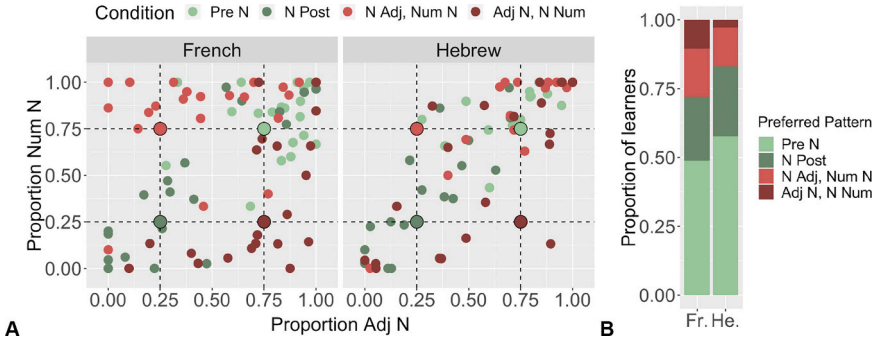

Fig. 10. A: Individual participant outcomes distributed across the space of possible 'grammars' in each condition for French- and Hebrew-speaking adult participants. B: Proportion of learners in each language group whose preferred pattern matches each of the four possible patterns.

two logistic mixed-effects regression models for each group. The first was a null model with only an intercept term. The second was a model including condition as a predictor. We then compared these models using likelihood ratio tests. For both groups, adding condition as a predictor led to a significant improvement to the fit of the model (French: $\chi^2 = 7.76$, $p = 0.05$; Hebrew: $\chi^2 = 22.86$, $p < 0.001$). To probe differences between conditions in each group further, all conditions were compared to each other. In the French-speaking group, participants in the Pre N condition produced marginally more of the dominant input than participants in both N Post ($\beta = 1.07 \pm 0.43$, $p = 0.07$) and Adj N, N Num ($\beta = 1.03 \pm 0.42$, $p = 0.07$), but did not differ from N Adj, Num N ($\beta = 0.84 \pm 0.41$, $p = 0.18$). No other differences approached significance (highest $\beta = 0.22 \pm 0.42$, lowest $p = 0.95$). In the Hebrew-speaking group, participants in the Pre N condition produced significantly more of the dominant input pattern than all other conditions (v. N Post: $\beta = 0.98 \pm 0.34$, $p = 0.02$; v. N Adj, Num N: $\beta = 1.50 \pm 0.35$, $p < 0.001$; v. Adj N, N Num: $\beta = 1.54 \pm 0.34$, $p < 0.001$). No other differences approached significance (highest $\beta = 0.56 \pm 0.32$, lowest $p = 0.30$).

3.2.3. *Preferred patterns*

As in Experiment 1, individual participants within each condition were quite variable in their behavior, with some shifting toward a non-input-like pattern. Fig. 10A illustrates the distribution of patterns

individual participants produced. Here we can see that across both groups, many participants shifted toward the (L2-like) harmonic pattern Pre N. This was most striking in the Hebrew group, where Fig. 10A shows that no learners regularized the L1-like non-harmonic pattern N Adj, Num N when that was the dominant input order they were trained on. Taking the two groups together, we can again classify each participants' preferred pattern based on the order they used most for each modifier type (see footnote 8). This yielded an overwhelming majority of harmonic choice (120/159), of which 82 were Pre N, compared to non-harmonic (36/159, of which 23 were L1-like N Adj, Num N), as confirmed by a two-tailed binomial test ($p < 0.001$).[14] Preferred patterns are summarized for each language group in Fig. 10B. It is worth noting that these shifts are qualitatively different from those we see in the English-speaking adult learners in Fig. 2 (Culbertson et al., 2012); no learners in that population shifted as radically toward a non-input-like corner of the space. For example, here we observe learners shifting from N Adj, Num N to *near-deterministic* Pre N or N Post, or even from Post N to Pre N. By contrast no learners in either group shifted from a harmonic input-pattern to a non-harmonic pattern (save perhaps one in the French group, who shifted from Pre N toward N Adj, Num N).

3.2.4. Consistency of preferred pattern use

As in Experiment 1, we also analyzed the consistency with which participants used their preferred patterns. This is summarized in Fig. 11. For both groups of participants, we used mixed-effects logistic regression models and model comparison to test whether adding preferred pattern as a predictor significantly improved the fit of the model. This was the case for Hebrew ($\chi^2 = 14.10$, $p = 0.003$), but not French ($\chi^2 = 5.18$, $p = 0.16$). To probe differences between conditions for the Hebrew group, all conditions were compared to each other (Adj N, N Num was again removed, since only 3 Hebrew participants chose this as their preferred pattern). Pre N differed significantly from N Adj, Num N ($\beta = 1.58 \pm 0.45$, $p = 0.002$), N Post differed marginally from N Adj, Num N ($\beta = 1.16 \pm 0.49$, $p = 0.08$), and Pre N and N Post did not differ from one another ($\beta = 0.26 \pm 0.33$, lowest $p = 0.86$). Note that the pattern appears to be qualitatively similar in French.

3.3. Discussion

In this experiment, we trained adult speakers of two non-harmonic languages–French and Hebrew–on one of four input patterns. Two patterns were non-harmonic, like their L1s, and two were harmonic. Importantly, one of the harmonic patterns was like English, in which almost all participants were bilingual. As in Experiment 1, this experiment provides evidence against the idea that participants' L1 pattern type drives their learning behavior. Rather, for both groups harmonic patterns fare better on almost all measures. Pre N was regularized most, and chosen most often as learners' preferred pattern. N Post was just as likely to be learners' preferred pattern as the L1-like N Adj, Num N, and those learners used it with a level of consistency that matched Pre N. This suggests that, as for children in Experiment 1, there is general harmony bias at work. Indeed, for Hebrew speaking adults, the preference for harmonic patterns over non-harmonic patterns is (qualitatively) almost as strong as that found in Hebrew-speaking children in Experiment 1. It is worth noting here that the difference observed between the two language groups in Experiment 1 is also found here. In both cases, French speakers were more likely to match their L1 pattern than Hebrew speakers. As noted above, it is possible that this reflects French speakers' experience with lexically-based restrictions on adjective order.

However, it is not obvious that a bias for harmony alone is driving the results in this case. In particular, there is reason to suspect that adult participants' L2 English experience may be playing a role. In particular, many learners shifted dramatically toward the English-like Pre N pattern, regardless of the dominant pattern they were trained on. This is unlike the behavior of English-speaking adults in Culbertson et al. (2012), who generally regularized their input pattern or shifted more subtly toward another. To explore whether this behavior is related to L1 experience, we looked at whether participants with a higher L2 proficiency score tended to (more strongly) prefer Pre N order. This is shown in Fig. 12, where no obvious pattern can be detected: learners of different proficiency levels are distributed evenly across the space. This is confirmed by a comparison of two linear regression models, which shows that adding English proficiency score does not improve predictions of how frequently learners use the English-like pattern ($\chi^2 = 1.82$, $p = 0.18$).[15] However, while there is no evidence that level of proficiency correlates with likelihood of producing an English-like pattern, it could still be that *any* substantial experience with English leads learners to do so.

To investigate further what is driving our results, we need a matched population with the *opposite* profile: English speakers who have substantial experience with a non-harmonic L2. If harmony facilitates language learning and use, one may also expect the harmony bias to manifest in the extent to which properties of the L2 influence the learning of a new language. More specifically, the L2 influence argued to account for French and Hebrew-speaking adults' Pre N preference in Experiment 2 may be reduced in Experiment 3 because participants' L2 is non-harmonic. If these learners exhibit less dramatic shifting toward their L2 than we saw in Experiment 2, then we will have additional evidence for the role of a general harmony bias in these bilingual adult populations. Specifically, such a result would suggest that degree of transfer of an L2 is more likely when the L2 pattern is preferred on learnability grounds.

4. Experiment 3: adult learners with L1 English, L2 non-harmonic

In Experiment 3 we test a third group of bilingual adults: English native speakers who are bilingual in a non-harmonic language (e.g., French or Spanish).

4.1. Method

4.1.1. Participants

Native English-speaking participants were 59 adults, recruited from the student population at the University of Edinburgh.[16] All participants completed a version of the LEAP-Q language background questionnaire (Marian et al., 2007). A histogram showing the distribution of composite scores for proficiency in their non-harmonic language is shown in Fig. 13.[17] This population is qualitatively well-matched to both the French and Hebrew speakers in terms of proficiency in their

[14] The remaining three participants could not be so classified (i.e., productions featured exactly 50% of each order for one or both modifier types). Two additional participants did not produce any usable phrases for one type of modifier, therefore they are not included in this analysis, nor in Figure 10 below.

[15] Running separate models for the two languages does not change this (French: $\chi^2 = 0.65$, $p = 0.42$; Hebrew: $\chi^2 = 1.70$, $p = 0.20$). Relatedly, the difference between the French and the Hebrew groups noted above could in principle reflect the fact that French speakers' English proficiency scores were slightly but significantly lower than Hebrew speakers' ($\beta = 0.75 \pm 0.26$, $p = 0.004$). This might have led to a reduced influence of the L2 pattern (e.g., relative to the L1 pattern). However, the lack of straightforward relationship between participants' proficiency with English and their use of English-like order in the experiment suggests this is not the case.

[16] One additional participant (in the Pre N condition) failed to produce any coherent responses in the production phase, and was therefore excluded from analysis.

[17] If a participant reported speaking more than one non-harmonic language, we used proficiency reports from the one they indicated as most dominant.

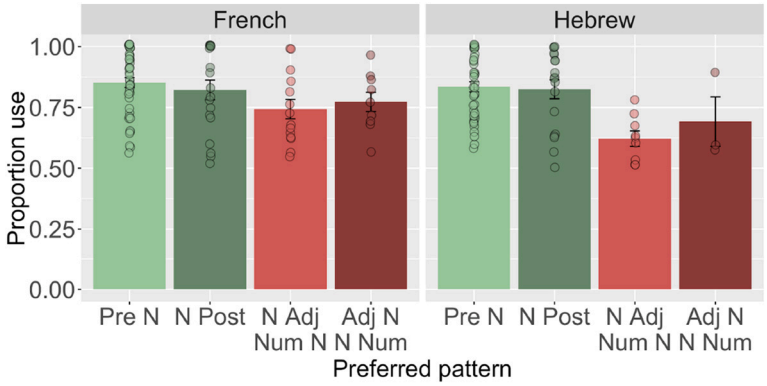

Fig. 11. Proportion use of each preferred pattern for French- and Hebrew-speaking adults. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means.

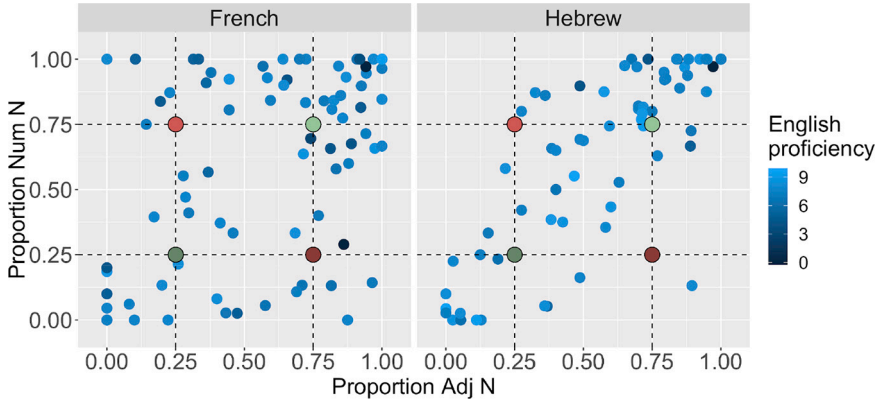

Fig. 12. French- and Hebrew-speaking adult outcomes colored according to self-reported English proficiency (composite score).

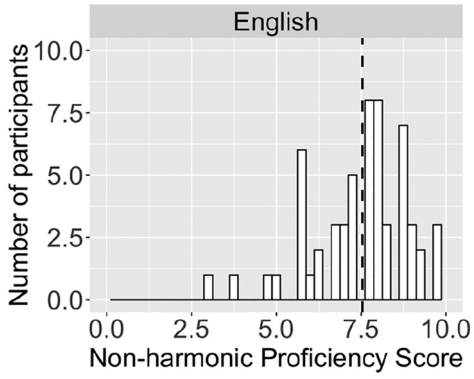

Fig. 13. Distribution of non-harmonic language composite proficiency scores for English-speaking (bilingual) adults. Dashed line shows the group mean.

non-L1 of interest. There was no significant difference between their non-harmonic language proficiency and the English proficiency scores of Hebrew participants ($\beta = 0.01 \pm 0.27$, $p = 0.98$). Like the Hebrew scores these were slightly but significantly higher than the French-speakers' English proficiency scores ($\beta = 0.74 \pm 0.27$, $p = 0.006$).

4.1.2. Stimuli

As in Experiment 2, participants were taught a language with 10 nouns and 10 modifiers (5 adjectives and 5 numerals). All lexical items were fully nonce. Nouns were two- or three-syllable words that consistently ended in 'a'. Modifiers were single syllable words. All lexical items are shown in Table 5. The visual stimuli were as in Experiment 2.

4.1.3. Procedure

The experimental procedure was identical to Experiment 2.

4.2. Results

4.2.1. Coding and vocabulary accuracy

Participants' productions in the critical test phase were coded for vocabulary accuracy and order (pre- or post-nominal) by a native speaker of English. Fig. 14 shows vocabulary accuracy across conditions. To determine whether there were significant differences across conditions in vocabulary accuracy, we fit two logistic mixed-effects regression models to the data. The first was a null model with only an intercept term. The second was a model including condition as a predictor. We then compared these models using likelihood ratio tests. Adding condition as a predictor did not significantly improve the fit of the model ($\chi^2 = 1.10$, $p = 0.78$).

4.2.2. Regularization of dominant input order

Fig. 14B shows average production of the dominant order across conditions.[18] As this figure suggests, for both groups, only participants whose dominant input order was N Post or N Adj, Num N regularized above the input level (N Post: $t(14) = 2.14$, $p = 0.05$; N Adj, Num N: $t(14) = 3.42$, $p = 0.004$; use of Adj N, N Num was significantly below

[18] As for Experiment 2, following Culbertson et al. (2012), we discarded trials in which the vocabulary was incorrect. This resulted in the exclusion of approximately 25% of trials. In almost all cases, participants produced two-word utterances (< 5% involved a missing word), therefore incorrect phrases involved either the noun or modifier (or both) produced incorrectly. However, note that including these trials does not have any substantial effect on the results reported here.

Table 5
IPA transcriptions (and meanings for adjectives and numerals) of English artificial language lexicon.

English					
Nouns		Adjectives		Numerals	
[gɹftə]	[slɛɹgemə]	[θɹæf]	'big'	[dof]	'two'
[nɛɹka]	[ræmpeɪzə]	[jɛv]	'small'	[kɛz]	'three'
[flaɪmə]	[wapogə]	[fuʃ]	'green'	[glaʊb]	'four'
[maʊgə]	[trəfundə]	[giʒ]	'blue'	[zadʒ]	'five'
[blɪfonə]	[powaɹtə]	[tʃɛɹg]	'furry'	[vɔɪtʃ]	'six'

4.3. Discussion

Taken together, these results suggest that adult English speakers bilingual in a non-harmonic language have been influenced to some degree by their L2. Results from the largely monolingual English-speaking adult population tested in Culbertson et al. (2012) revealed a clear bias for the two harmonic patterns: participants regularized significantly more in the two harmonic conditions than in the non-harmonic N Adj, Num N condition, and they generally shifted toward harmonic patterns (see Figs. 1A and 2A). By contrast, in this bilingual population, there was no difference in regularization of the non-harmonic L2 order, N Adj, Num N compared to either of the two harmonic orders. Likewise, these two orders were equally likely to be used as participants' preferred patterns. They tended to shift the other Pre N and Adj N, N Num in the direction of increasingly post-nominal adjectives. It's worth noting, however, that there is a high degree of variation in participants' behavior both here and in the monolingual population reported in Culbertson et al. (2012). Further, mixed-effects logistic regression models assessing regularization by participants trained on majority N Adj, Num N across the two studies revealed only a marginally significant difference ($\chi^2 = 3.26$, $p = 0.07$). This suggests that the influence of the L2 in this case is relatively weak. Importantly, as in Experiments 1 and 2, there is clearly no advantage for the alternative non-harmonic pattern, suggesting that abstract transfer of pattern-type is not the explanation for learners' shifts across these experiments.

Notably, despite the apparent influence of these learners' L2, N Adj, Num N was not the magnet for shifting and regularization that Pre N was for learners in Experiment 2. If the preference for harmony seen in Experiment 2 was due to L2 transfer alone, then we would expect English learners in Experiment 3 to have shifted toward non-harmonic, or at least N Adj, Num N, with similar strength to the shift toward the harmonic English pattern manifested by French and Hebrew speakers in Experiment 2. To assess the possibility that L2 influence on learning is modulated by whether L2 has a harmonic pattern or not, we compared the number of participants who used an L2 vs. non-L2 pattern in both experiments. In Experiment 2, 82 participants (53%) were classified as using the L2 pattern (Pre N), and 74 (47%) were classified as using another pattern; in Experiment 3, 19 participants (33%) were classified as using the L2 pattern (N Adj, Num N), and 39 (67%) were classified as using another pattern. This difference was significant, confirming that a harmonic L2 pattern served as a stronger attractor for learners than a non-harmonic L2 ($\chi^2 = 5.88$, $p = 0.02$).

As in Experiment 2, we also explored the relationship between participants' proficiency with a non-harmonic language and their use of an L2-like N Adj, Num N order in the experiment. This is shown in Fig. 17. Again, learners of different proficiency levels are distributed evenly across the space. This is confirmed by a comparison of two linear regression models, which shows that adding the non-harmonic

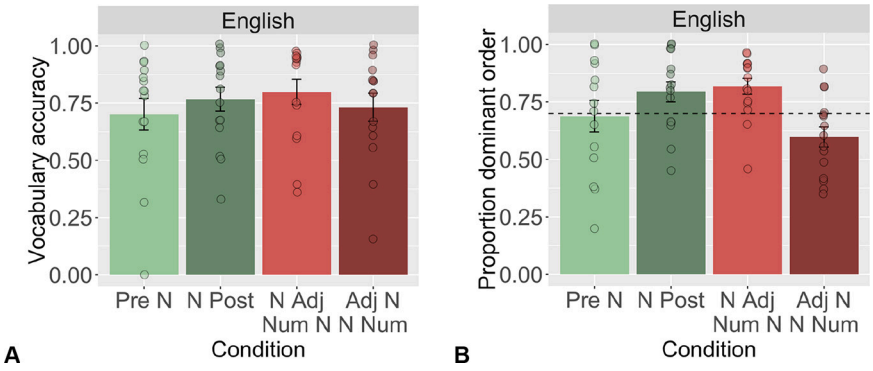

Fig. 14. A: Vocabulary accuracy during critical testing phase across conditions for English-speaking (bilingual) adult participants. B: Proportion use of the dominant order in each condition (dotted line shows the frequency of the dominant pattern in the input). Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means. Bars show group average, points show individual participants (jittered), error bars represent standard error on by-participant means.

the input: $t(14) = -2.36$, $p = 0.03$). To determine whether there were significant differences across conditions in use of the dominant input order, we fit two logistic mixed effects regression models to the data. The first was a null model with only an intercept term. The second was a model including condition as a predictor. We then compared these models using likelihood ratio tests. Adding condition as a predictor led to a significant improvement to the fit of the model ($\chi^2 = 11.27$, $p = 0.01$). To probe differences between conditions further, all conditions were compared to each other. There was a significant difference between the two non-harmonic conditions, with participants in the L2-like N Adj, Num N condition using input order more often than participants in Adj N, N Num ($\beta = 1.29 \pm 0.43$, $p = 0.01$). There was also a significant difference in the same direction between N Post and Adj N, N Num ($\beta = 1.30 \pm 0.43$, $p = 0.01$). No other differences approached significance (highest $\beta = 0.65 \pm 0.44$, lowest $p = 0.44$).

4.2.3. Preferred pattern use

Fig. 15A illustrates the distribution of patterns individual participants produced. Classifying the pattern each participant used most (in more than 50% of their productions) yielded a slim majority of harmonic choice (33/59), of which 21 were Post N), compared to non-harmonic (25/59, of which 19 were L2-like N Adj, Num N), this was not significant according to a two-tailed binomial test ($p = 0.43$).[19] This is summarized in terms of the proportion of learners with each preferred pattern in Fig. 15B.

4.2.4. Consistency of preferred pattern use

Finally, consistency of preferred pattern use was analyzed as in Experiments 1 and 2. This is shown in Fig. 16. Mixed-effects logistic regression models and model comparison indicated that adding preferred pattern as a predictor did not significantly improve the fit of the model ($\chi^2 = 2.37$, $p = 0.50$).

[19] The remaining participant could not be so classified (i.e., productions featured exactly 50% of each order for one or both modifier types).

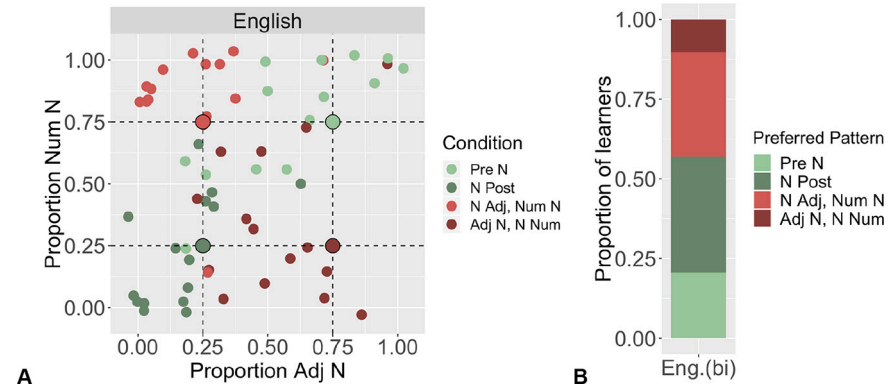

Fig. 15. A: Individual participant outcomes distributed across the space of possible 'grammars' in each condition for English-speaking (bilingual) adult participants (jittered). B: Proportion of learners whose preferred pattern matches each of the four possible patterns.

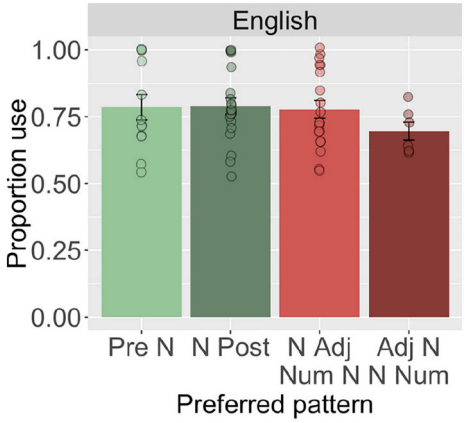

Fig. 16. Proportion use of each preferred pattern for English-speaking (bilingual) adults. Bars show group averages, points show individual participants (jittered), error bars represent standard error on by-participant means.

language proficiency score does not improve predictions of how frequently learners use the L2-like pattern ($\chi^2 = 0.47$, $p = 0.50$).

5. General discussion

The frequency of harmonic word order patterns across languages has long been noted by linguists, who have generated a number of possible explanations for it. Here we have identified two general classes of explanation which differ critically in the role they ascribe to the human cognitive system. In particular, it has been argued that harmonic patterns are common among the world's languages because they are easier to learn and use than non-harmonic patterns (e.g., Culbertson & Kirby, 2016; Hawkins, 2004; Pater, 2011; Vennemann, 1976). However, there is also evidence from the historical record showing that some harmony patterns may arise because of shared history. For example, there is a very strong tendency for languages which put verbs before their dependent objects to also put adpositions before their dependent nouns, and vice versa. This may be because verbs are a common source of new adpositions, therefore these adpositions naturally share the order of the verbs that served as their source (Givón, 1975). If harmony is generally the result of diachronic processes, then there is no need to posit an independent cognitive preference for harmonic order (Aristar, 1991; Kaufman, 2009). Indeed, this is part of a more general point made by many researchers who advocate the view that diachronic processes, and not individual-level cognitive factors, drive typological trends (e.g., Blevins, 2004; Bybee, 2006, 2008; Cristofaro, 2017; LaPolla, 2010; Ohala, 1993).

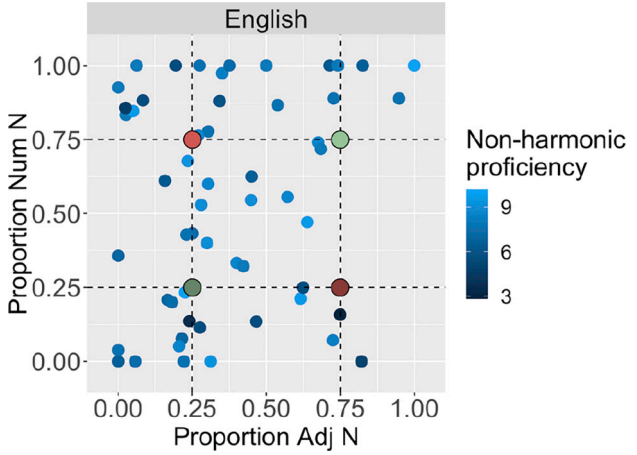

Fig. 17. English-speaking (bilingual) outcomes colored according to self-reported non-harmonic language proficiency (composite score).

While debate about which of these general explanations is the right one has been ongoing since the 1970's, recent work has used artificial language learning experiments to look for independent behavioral evidence that the cognitive system has a role to play in shaping typology. Culbertson et al. (2012) and Culbertson and Newport (2015) tested whether English-speaking adults and children prefer harmonic orders of nouns with different types of modifiers. They taught learners patterns of variable order which tended to be either harmonic or non-harmonic. For example, in one harmonic input language post-nominal adjectives were used most of the time, and so were post-nominal numerals. In a non-harmonic input language, adjectives were mostly *post*-nominal but numerals were mostly *pre*-nominal. They found that adults were more likely to regularize harmonic patterns they were trained on, and children often produced harmonic outputs even when the input was mainly non-harmonic (see also Culbertson & Newport, 2017). Importantly, they treated pre-nominal harmony and post-nominal harmony the same–that is, they did not show any special preference for their native language order over the opposite. These experiments are consistent with a hypothesized link between human learning and word order harmony. However, English-speakers' preferences in this task could also have reflected abstract transfer of the *type* of pattern found in English. For example, English speakers might be used to treating adjectives and numerals the same, and therefore any pattern which orders them similarly may have a learning advantage. This kind of abstract transfer was suggested as an explanation by Goldberg (2013).

5.1. New evidence from cross-linguistic and multilingual populations

In order to rule out this explanation, and provide further evidence for a harmony bias in second language learning, we have tested a

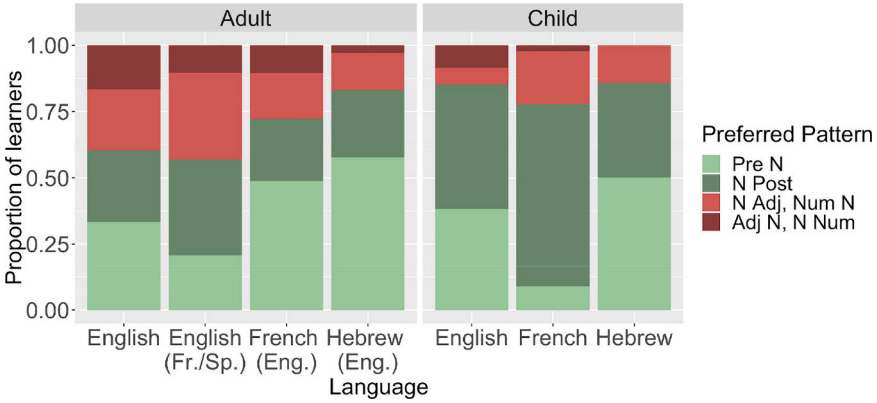

Fig. 18. Summary of all data by age and L1 (with L2's in parentheses). English-speaking adult data are from Culbertson et al. (2012); English-speaking child data are from Culbertson and Newport (2015). Y-axis shows the proportion of participants in each group whose preferred pattern (used in > 50% of utterances) matched each of the four possible patterns. Green patterns (bottom two in each bar) are harmonic, red are non-harmonic.

number of populations who have substantial or exclusive experience with a non-harmonic language. Our main aim was to assess whether abstract transfer from the L1 could explain previous findings, or whether despite their prior linguistic experience, these learners nevertheless prefer harmony. We also explored whether and how L2 experience might impact word order learning. To summarize our results here, we focus first on learners' *preferred patterns*–calculated based on participants' most frequently produced orders for adjectives and numerals (e.g., a participant who produced > 50% N Adj *and* > 50% N Num is classified as preferring N Post). Fig. 18 shows the proportion of participants in each group who predominantly used each of the four patterns of interest. This is a summary of our main results. We have also included the same data from monolingual English speakers (Culbertson et al., 2012; Culbertson & Newport, 2015). The main take-away from this analysis is that across all populations, preferred patterns were more likely to be harmonic than non-harmonic (i.e., there is more green than red). This result is most clear in the child populations we tested: monolingual speakers of both harmonic and non-harmonic languages preferred to use harmonic patterns. This preference can also be seen, albeit more subtly, across the adult populations tested: regardless of their L1, harmonic patterns enjoyed an advantage. These results therefore show that abstract transfer of the L1 pattern type cannot explain the previously reported data on English speakers. Rather, a harmony bias impacts learning even in the face of substantial experience with a non-harmonic language.

Interestingly, participants' L2 experience did have an impact on the patterns they learned, and this interacted with the harmony bias. In particular, French- and Hebrew-speaking adult learners who were L2-speakers of English showed a clear preference for English-like prenominal harmony over the other patterns. Indeed, many participants used this pattern near-deterministically, even when their input did not resemble it. Importantly though, in both populations the post-nominal harmonic pattern was just as likely to be used as the native-like non-harmonic N Adj, Num N. There was also some evidence that L2 experience with a non-harmonic (N-Adj, Num N) language led English learners to use that order more often than a comparable population of English monolinguals. However, crucially, across the two types of bilingual populations we tested, learners were significantly more likely to choose the L2 order *if it was harmonic*.[20] This suggests that the harmony bias impacted the degree of L2 transfer. We return to this below. However, first we discuss the harmony bias in more general terms.

5.2. Harmony as simplicity

Taken together with the results of Culbertson et al. (2012) and Culbertson and Newport (2015, 2017), these studies provide evidence of the experience-independent nature of the harmony bias. Such a bias is consistent with accounts of harmony in which the human cognitive system plays a role. For example, a number of researchers have argued that the harmony bias is a reflex of a more general cognitive bias for simplicity (Culbertson & Kirby, 2016; Culbertson & Newport, 2015, 2017). Intuitively, a grammar which encodes distinct ordering rules for each type of modifier (or dependent) is more complex than one which has a single general rule for ordering heads and dependents. Put another way, learners can generalize across head-dependent pairs in a harmonic language, but not in a non-harmonic one. Of course, this does not mean that processes of historical change do not also play a role in shaping language typology. It simply suggests that cognition is a part of the picture. It is also worth noting that these experiments target the nominal domain, while work on the role of language change has provided the strongest evidence for true cross-category harmony between the verb phrase and the adpositional phrase (e.g., Aristar, 1991; Givón, 1975). Indeed, the cross-linguistic tendency for harmony in that domain is stronger than in the nominal domain, suggesting that when a cognitive bias aligns with a strong diachronic pathway, this is reflected in the typology. Future work should investigate whether the cognitive bias found here for ordering within the noun phrase applies to clear cases of cross-category harmony like this. Particularly important, for example, would be cases in which the grammatical categories involved are distinct; here, and in cross-category harmony between the verb phrase and the adpositional phrase, harmony can in principle be reformulated as a preference to align nouns (rather than heads) across phrase types.

While the proposal we have described above is that harmony is a case of a more general bias for simplicity in learning, this is not the only cognitive mechanism that could drive harmony. As mentioned in the introduction, a number of authors have argued that at least some cases of harmony could result in part from processing pressures (e.g., Hahn et al., 2018, Hawkins, 2009). For example, in some cases non-harmonic patterns involve longer dependencies, which might introduce heavier demands on working memory. However, this applies only to cases where multiple phrase types co-occur in a single utterance (e.g., `kick the can in the road' or in our case 'two black cats'). These types of complex phrases are absent from our task, and therefore are unlikely to be driving participants' preference for harmony. Another possible

[20] As a reviewer points out, there is also more variation in the behavior of the three adult populations tested here relative to the original English-speaking adult population tested in Culbertson et al. (2012). It remains an open question exactly what drives this, since the age and education levels of all populations were similar. Readers familiar with the latter work may recall that in that experiment, participants were given online feedback on their responses during production, which might have had the effect of restricting behavior. However, the result (with similar levels of variation) was replicated without feedback in Culbertson and Smolensky (2012). An alternative possibility is that multilingualism leads to this increase in variance; participants in the populations tested here may simply be influenced by more and more varied linguistic knowledge than monolingual English speakers.

(*footnote continued*)

mechanism comes from production priming (e.g., see Bock & Griffin, 2000); it could be that harmonic word orders are preferred because they allow the producer to re-use the same kind of structure as was used in previous utterances. Interestingly, this kind of mechanism has been proposed as an explanation for regularization as well (Ferdinand et al., 2019).

It is also worth noting that despite the bias for harmony shown in this experimental setting, any such bias can be overridden; natural languages productively use non-harmonic noun phrase word orders, and these are sustained over generations of learners (i.e. from Latin to the Romance languages in the present day). Indeed, data on acquisition of noun phrase word order by children suggests that it is mastered very early (Cipriani et al., 1993, Montrul, 2004, Prévost, 2009, also see footnote 9). How then, would a weak harmony bias play a role in shaping typology? This is a critical question to ask given that artificial language learning experiments are an extremely simplified analogue to natural language learning (either by children or adults). One possibility is that, rather than preventing learners from acquiring non-harmonic noun phrase word orders, the bias encourages change to non-harmonic systems when an opportunity arises. For example, change of this kind might be likely to arise when there is variation in the system (e.g., due to other processes of historical change), or when contact between two language populations occurs. Our results make the clear prediction that, all things equal, if speakers of a language with harmonic noun phrase order come into contact with speakers of a language with a non-harmonic order, the influence should be asymmetrical. This would be in line with research on contact-induced language change which has generally been argued to lead to simplification (e.g., see Miestamo et al., 2008, though this is not always the case, e.g., see discussion in Meakins et al., 2019). Another possible context for the harmony bias to play a role relates directly to Experiments 2 and 3; it might be the case that bilinguals using both a harmonic and non-harmonic language could introduce harmonizing “errors” into a non-harmonic language more often than the reverse. Interestingly, there is some evidence for this process from data on bilingual acquisition: several studies report that children bilingual in a Romance and Germanic language produce more non-adult-like adjective orders in their Romance language compared to their Germanic language, and more such reversals than monolinguals (e.g., see Nicoladis, 2006; Rizzi et al., 2013).

5.3. Nominal typology and N-initial patterns

Returning to the typology of nominal order (shown in Table 1B), it is worth noting that in addition to a tendency for harmonic patterns to outnumber non-harmonic patterns, there is also a tendency for post-nominal adjectives to outnumber pre-nominal adjectives. This can be seen when comparing *within* pattern type: N Post pattern is the most common harmonic pattern (over Pre N) and N-Adj, Num N is the most common non-harmonic pattern (over Adj N, N Num). The results across experiments shown in Fig. 18 suggest that the typological preference for N Adj may also reflect a cognitive bias (independent from harmony). First, despite substantial experience with non-harmonic patterns (as L1 or L2) very few participants in our experiments regularized Adj N, N Num when trained on a language in which this order was dominant, and few used it as their most frequent pattern (e.g., 20/308 across the experiments reported here). This cannot be explained purely by the fact that no participants tested had that pattern as their L1 or L2; Post N was readily used (105/308 across the experiments reported here); French-speaking children strongly preferred this pattern; English-speaking children were numerically more likely to use it. Why might post-nominal adjectives be preferred by learners?

One possibility is that it stems from the semantics of adjectives: in many cases, the meaning of an adjective depends on the noun it modifies (i.e., in gradable adjectives like ‘tall’ or adjectives like ‘skilled’, see e.g., Kamp & Partee, 1995). If the noun comes first, then the adjective can be interpreted immediately. Recent experimental work has also shown that speakers use adjectives differently depending on whether they are pre- or post-nominal. For example, Rubio-Fernandez et al. (2018) finds that speakers of a language with pre-nominal adjectives (English) redundantly use color adjectives more often than speakers of a language with post-nominal adjectives (Spanish). For example, English speakers might say ‘the red triangle’ to prompt a conversation partner to pick out a triangle in a scene with no competing triangle present (i.e., a scene with a red triangle and a blue circle). She argues that this is due to the fact that in pre-nominal adjective languages, using an adjective, even redundantly, can facilitate communication (i.e., allow a listener to identify a referent sooner). In post-nominal languages, redundant adjectives are less likely to facilitate communication. While this is not direct evidence that having the noun before the adjective is better, it does suggest that if nouns are generally better descriptors of referents, then having them first would be an advantage. Alternatively, if nouns are generally easier to access, then speakers may tend to produce them first (e.g., Fukumura, 2018).[21] While independent evidence along these lines would help us to understand *why* post-nominal adjectives may be more common, results from the experiments reported here suggest that whatever bias is at play can affect word order learning. To summarize, the typology of nominal word order suggests two independent biases: a harmony bias and a preference for post-nominal adjectives. Our results are compatible with the hypothesis that both are at play during learning: harmonic orders are generally preferred to non-harmonic orders, and despite neither being the L1 or L2 pattern of our participants, N Post is learned much more readily than Adj N, N Num.

5.4. Implications for theories of Ln acquisition

Finally, although they were not designed to address this, our results may also speak to theories of the early stages of third-language or Ln learning. These theories are formulated to explain patterns of facilitation or inhibition that previously learned languages have on the learning of a new language. Some theories argue that the typological or structural similarity of a new language to previously learned languages largely determines patterns of transfer (e.g., Rothman et al., 2011, Westergaard et al., 2017). For example, Rothman et al. (2011) argues that *phonological similarity* most strongly determines early transfer patterns. That is clearly not what happens in our studies; all input lexicons were designed to be phonologically plausible in participants' L1 (not their L2), and yet L1-like orders were not at an advantage. For example, L1 speakers of French and Hebrew clearly preferred using the order found in their L2 (English)–which does not match the phonology of the lexical items in the input. Theories which argue that *structural similarity* determines early transfer patterns would predict facilitation in our adult populations regardless of the input pattern, since they all know both a harmonic and a non-harmonic language (Westergaard et al., 2017). Again, this is not what we see. At least one prominent hypothesis is that the second language (as opposed to the L1) has a privileged status, regardless of its similarity to the new language (e.g., Bardel & Falk, 2007; Bardel & Falk, 2012). Our results are largely consistent with this, showing a clear influence of the L2 in all cases. What we would suggest is that these theories may also need to take into account the possibility that some patterns–regardless of whether they are found in the L1 or L2–are more difficult to learn than others. In other words, the degree of L2 influence may be modulated by general learnability, with more easily learnable patterns more likely to influence the bilingual speaker. Of course, here we have only focused on one particular contrast – between harmonic and non-harmonic orders in the

[21] A reviewer suggests the possibility that the noun may be more pragmatically important, and therefore produced first, given that in our task neither the object nor the modifier need to be used contrastively (our production trials involve a single picture only).

noun phrase – and have only compared two populations. Therefore, additional work would be needed to confirm the generalizability of this claim.

On the other hand, the idea that both L1 and L2 experience may influence Ln learning in the lab–likely not a surprise to researchers studying Ln acquisition–is not always acknowledged in the artificial language learning literature. Our results suggest there is a complex interplay between L1, L2, and cognitive biases which needs to be taken into account when interpreting these types of experiments, particularly with adults.

6. Conclusion

Word order harmony is one of the most well-known and well-studied typological universals–at least from a descriptive and theoretical perspective. However, there is long-standing disagreement about what role, if any, the cognitive system plays in driving the tendency for harmony. Recently, artificial language learning experiments have been used to provide an independent source of evidence for the role of human cognition in driving harmony. A series of studies reported that harmonic noun phrase orders were preferred over non-harmonic orders by English-speaking adults and children (Culbertson et al., 2012; Culbertson & Newport, 2015, 2017). However, this evidence is potentially problematic, given that English is itself a harmonic language (Goldberg, 2013). Here we tested whether the harmony bias found in these experiments was due to abstract transfer from the L1–a preference for harmonic pattern types, driven by experience with a harmonic L1. We did this by testing learners who have substantial experience with a *non-harmonic* language. An abstract transfer account predicts that these learners should prefer non-harmonic patterns. What we found instead was evidence for a bias in favor of harmony across all populations. Although in some cases this bias interacted with prior language experience in complex ways, our results support the hypothesis that a cognitive bias for harmony may have shaped language typology.

CRediT authorship contribution statement

Jennifer Culbertson: Conceptualization, Methodology, Software, Formal analysis, Visualization, Data curation, Writing - original draft. **Julie Franck:** Conceptualization, Methodology, Writing - review & editing. **Guillaume Braquet:** Methodology, Investigation. **Magda Barrera Navarro:** Methodology, Investigation. **Inbal Arnon:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgments

The authors would like to thank Samuel Schmid, Amir Efrati, Hila Merchav, Yuval Braeman as well as all participants who took part in our studies. This work was supported by the Economic and Social Research Council [grant number ES/N018389/1] and by the European Research Council (ERC) under the European Union's Horizon 2020 - Research and Innovation Framework Programme (grant agreement no 757643).

Supplementary material

All data reported in this paper is available here: https://osf.io/fdkh3/.

References

Aristar, A. R. (1991). On diachronic sources and synchronic pattern: An investigation into the origin of linguistic universals. *Language, 67*, 1–33.
Baker, M. (2001). *The atoms of language: The mind's hidden rules of grammar.* New York, NY: Basic Books.
Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of germanic syntax. *Second Language Research, 23*, 459–484.
Bardel, C., & Falk, Y. (2012). Behind the L2 status factor: A neurolinguistic framework for L3 research. *Third language acquisition in adulthood* (pp. 61–78). .
Bates, D. (2010). lme4: Mixed-effects modeling with R. http://lme4.r-forge.r-project.org/book.
Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns.* New York: Cambridge University Press.
Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*(2), 177.
Braquet, G., & Culbertson, J. (2017). Harmony in a non-harmonic language: Word order learning in French children. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711–733.
Bybee, J. (2008). Formal universals as emergent phenomena: The origins of structure preservation. In J. Good (Ed.). *Linguistic universals and language change* (pp. 108–121). New York: Oxford University Press.
Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures.* Cambridge, MA: MIT Press.
Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry, 36*(3), 315–332.
Cipriani, P., Chilosi, A.-M., Bottari, P., & Pfanner, L. (1993). *L'acquisizione della morfosintassi in italiano.* Padova: Unipress.
Cristofaro, S. (2017). Implicational universals and dependencies. In N. Enfield (Ed.). *Studies in diversity linguistics*. (pp. 9–22). Language Science Press.
Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass, 6*(5), 310–329.
Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain general biases have domain specific effects. *Frontiers in Psychology, 6*.
Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition, 139*, 71–82.
Culbertson, J., & Newport, E. L. (2017). Innovation of word order harmony across development. *Open Mind: Discoveries in Cognitive Science, 1*, 91–100.
Culbertson, J., & Smolensky, P. (2012). A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science, 36*, 1468–1498.
Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*, 306–329.
Culbertson, J., Smolensky, P., & Wilson, C. (2013). Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in Cognitive Science, 5*, 392–424.
Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language, 35*, 99–127.
Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.
Dunn, M., Greenhill, S., Levinson, S., & Gray, R. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature, 473*, 79–82.
Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(05), 429–448.
Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition, 184*, 53–68.
Fox, G., & Thuilier, J. (2012). Predicting the position of attributive adjectives in the French NP. *New directions in logic, language and computation* (pp. 1–15). Springer.
Fukumura, K. (2018). Ordering adjectives in referential communication. *Journal of Memory and Language, 101*, 37–50.
Givón, T. (1975). Serial verbs and syntactic change: Niger-Congo. In C. Li (Ed.). *Word order and word order change* (pp. 47–112). Austin, TX: University of Texas Press.
Givón, T. (1979). *On understanding grammar.* New York: Academic Press.
Givón, T. (1984). Universals of discourse structure and second language acquisition. In W. E. Rutherford (Ed.). *Language universals and second language acquisition* (pp. 109–136). Amsterdam: John Benjamins.
Goldberg, A. E. (2013). Substantive learning bias or an effect of familiarity? Comment on Culbertson, Smolensky, and Legendre (2012). *Cognition, 127*, 420–426.
Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.). *Universals of language* (pp. 73–113). Cambridge, MA: MIT Press.
Hahn, M., Degen, J., Goodman, N. D., Jurafsky, D., & Futrell, R. (2018). An information-theoretic explanation of adjective ordering preferences. *Proceedings of the 40th annual conference of the Cognitive Science Society*.
Harbour, D. (2016). *Impossible persons.* MIT Press.
Hawkins, J. A. (1979). Implicational universals as predictors of word order change. *Language, 55*, 618–648.
Hawkins, J. A. (2004). *Complexity and efficiency in grammars.* Oxford: Oxford University Press.
Hawkins, J. A. (2009). Language universals and the performance-grammar correspondence hypothesis. In M. H. Christiansen, C. Collins, & S. Edelman (Eds.). *Language universals* (pp. 54–78). Oxford University Press.
Hayes, B., Kirchner, R., & Steriade, D. (2004). *Phonetically based phonology.* Cambridge University Press.
Heine, B., & Kuteva, T. (2008). Constraints on contact-induced linguistic change. *Journal of Language Contact, 2*(1), 57–90.
Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346–363.
Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30–66.
Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57*, 129–191.
Kaufman, D. (2009). Austronesian nominalism and its consequences: A tagalog case study. *Theoretical Linguistics, 35*, 1–49.

Keenan, E. L. (1979). On surface form and logical form. *Studies in the Linguistic Sciences, 8*, 1–41.
LaPolla, R. J. (2010). *Problems of methodology and explanation in word order universals research.* Universitätsbibliothek Johann Christian Senckenberg.
Levinson, S. C., & Evans, N. (2010). Time for a sea-change in linguistics: Response to comments on "the myth of language universals". *Lingua, 120*(12), 2733–2758.
Mallinson, G., & Blake, B. J. (1981). *Language typology: Cross-linguistic studies in syntax.* Amsterdam: North-Holland.
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967.
McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., Bedini, C., Bertomeu Castell´o, N., & Jungmee, L. (2013). Universal dependency annotation for multilingual parsing. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers). Vol. 2. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers)* (pp. 92–97).
Meakins, F., Hua, X., Algy, C., & Bromham, L. (2019). Birth of a contact language did not favor simplification. *Language, 95*, 294–332.
Miestamo, M., Sinnemäki, K., & Karlsson, F. (2008). *Language complexity: Typology, contact, change. Vol. 94.* John Benjamins Publishing.
Montrul, S. A. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition.* John Benjamins.
Moreton, E., & Pater, J. (2012a). Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass, 6*(11), 686–701.
Moreton, E., & Pater, J. (2012b). Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass, 6*(11), 702–718.
Nicoladis, E. (2006). Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Bilingualism: Language and Cognition, 9*, 15–32.
Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.). *Historical linguistics: Problems and perspectives* (pp. 237–278). London: Longman.
Pater, J. (2011). Emergent systemic simplicity (and complexity). *McGill working papers in linguistics* (pp. 22). .
Prévost, P. (2009). *The acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition.* Philadelphia: John Benjamins.
R Development Core Team (2010). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing Vienna.
Rizzi, S., Gil, L. A., Repetto, V., Geveler, J., & Müller, N. (2013). Adjective placement in bilingual Romance-German and Romance-Romance children. *Studia Linguistica, 67*, 123–147.
Rothman, J., Iverson, M., Judy, T., & Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research, 27*, 107–127.
Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2018). *Why searching for a blue triangle is different in English than in Spanish.* PsyArXiv (October, 13).
Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition, 116*, 44–449.
Thomason, S. G. (2001). *Language contact.* Edinburgh University Press.
Travis, L. (1984). *Parameters and effects of word order variation.* (Ph.D. dissertation MIT).
Vennemann, T. (1976). Categorial grammar and the order of meaningful elements. *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday. Vol. 3. Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday* (pp. 615–634).
Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The linguistic proximity model. *International Journal of Bilingualism, 21*, 666–682.