Learners' harmonic preferences are modulated by lexical retrieval difficulty

Yiyun Zhao & Masha Fedzechkina[*]

Abstract. Typological work has established the existence of language universals – features or combinations of features that (co-) occur in unrelated languages more frequently than expected by chance. The origins of language universals are a fundamental question in language sciences as these universals are considered a reflection of cognitive mechanisms underlying human language. In this study, we use a miniature artificial language learning paradigm to explore whether a well-known universal – a preference for harmonic word ordering between adpositional and verb phrases (i.e., placing the head in a consistent position with respect to the complements across the two phrase types) – originates in biases during language learning and whether this preference interacts with memory constraints (operationalized as lexical retrieval difficulty). We first trained participants on miniature languages containing adpositional phrases of a deterministic word order (either prepositional or postpositional inputs) and then briefly exposed them to simple transitive sentences (verb-subject-object and subject-object-verb order, equally frequent in the input). At test, we asked learners to describe simple transitive scenes. We found that in the hard lexical retrieval condition, learners exposed to a postpositional language showed a preference for harmonic ordering but learners of the prepositional language did not, which is only partially consistent to the typological distribution. In the easy lexical retrieval condition, learners of neither the postpositional nor prepositional language showed a preference for harmonic ordering, indicating that this preference is modulated by memory constraints.

Keywords. harmonic ordering; statistical language universals; miniature artificial language learning; language evolution

1. Introduction. Natural languages vary in many aspects, but this variation is not random. Work in linguistic typology has established the existence of *statistical language universals* – features or combinations of features that (co)-occur more frequently than expected by chance in unrelated languages (e.g., Greenberg, 1963; Dryer, 1992). Ever since their discovery, understanding the nature of language universals has been a fundamental question in the language sciences as most researchers agree that it could shed light on the cognitive mechanisms of human language processing and acquisition (e.g., Chomsky, 1965; Christiansen & Charter, 2008; Hawkins, 2014).

One of the most well-known language universals is harmonic word ordering: Languages tend to place heads either consistently before or after their dependents (Greenberg, 1963). An example of such harmonic word order pattern is the correlation of head position between the adpositional phrase (AP) and verb phrase (VP) (Greenberg, 1963; Dryer, 1992; Hawkins, 1983) – head-initial (verb-object) languages tend to be prepositional and head-final (object-verb) languages tend to be postpositional. Harmonic ordering has also been found in the nominal domain

– languages tend to place the head (noun) either consistently before or after its dependents (adjectives, numerals, genitive, and relative clauses) (Greenberg, 1963; Hawkins, 1983).

Understanding the underlying causes of harmonic ordering preferences has been a central question in many linguistic theories. A long-standing view attributes this cross-linguistic tendency to individual biases in language learning or processing (e.g., Chomsky, 1965; Fodor, 2001; Christiansen & Charter, 2008; Hawkins, 2014). Views differ, however, as to the exact nature of these biases. One influential view explains harmonic ordering as a linguistic-specific constraint on grammar (e.g., Baker, 2001; Travis, 1984). Others argue that the preference for harmonic ordering instead reflects domain-general constraints such as a preference for simpler grammars (Culbertson & Kirby, 2016) or processing efficiency (Hawkins, 2014; Dryer, 1992; Ferrer i Cancho, 2004). Recent work, however, has called into question the cognitive roots of the preference for harmonic ordering, thus posing a problem for both domain-general and domain-specific explanations of this preference. For instance, Dunn, Greenhill, Levinson & Gray (2011) argued that the prevalence of harmonic orders is explained by a shared linguistic phylogenetic relationship among languages and Whitman (2008) proposed that harmonic ordering may result from a process of grammaticalization (adpositions may have originated from verbs that underwent the process of grammaticalization, therefore resulting in the harmonic word orders across the AP and VP). Thus, despite the large amount of attention devoted to harmonic ordering, its underlying causes remain unclear.

One potential reason for the lack of consensus on the origin of harmonic preferences is the inherent limitations of typological data (a traditional data source in the study of language universals). First, typological data suffers from the sparsity of independent data points as many languages have shared ancestors or a history of prolonged contact with each other (e.g., Dunn et al., 2011; Cysouw, 2010; Jaeger, Graff, Croft & Pontillo, 2011; Rafferty, Griffiths & Klein, 2014). Second, and more importantly, it is difficult to draw conclusions about the underlying causes of language universals based on typological data.

In light of these limitations, an innovative experimental approach –*miniature artificial language learning* – has been recently gaining popularity as a complement to typological data (Hudson Kam & Newport, 2005; Kirby, Cornish & Smith, 2008; for a discussion see Fedzechkina, Newport & Jaeger, 2016). In these experiments, researchers study how participants learn novel miniature languages designed in the laboratory. This paradigm does not suffer from data sparsity as any number of languages can be generated to study specific questions about language universals. More importantly, miniature language learning can directly probe the underlying causes of language universals. To study this, researchers design miniature languages to contain variable input (similar to a pidgin) and investigate whether and how learners change the inconsistent input after they have successfully acquired the miniature language. A common finding in this line of work is that learners prefer cross-linguistically frequent patterns over the less frequent alternatives at all levels of linguistic organization: phonology (Finley & Badecker, 2012; Wilson, 2006), morphology (Fedzechkina, Jaeger & Newport, 2012; Hudson Kam & Newport, 2009, 2005), and syntax (Culbertson & Newport, 2015; Culbertson, Smolensky &Legendre, 2012; Fedzechkina, Chu &Jaeger, 2018; Tily, Frank & Jaeger, 2011).

Miniature artificial language learning has been applied to study harmonic ordering. For instance, Culbertson, Smolensky & Legendre (2012) found a preference for harmonic ordering in the nominal domain. They exposed participants to miniature languages that ordered adjectives and numerals inconsistently with respect to the noun. They found that adult learners preferred using harmonic orders (i.e., they preferred to use both adjectives and numerals either pre-

nominally or post-nominally) and a non-harmonic but typologically frequent numeral-noun/noun-adjective order. Later work has further shown that child learners have an even stronger preference for harmonic ordering compared to adults (Culbertson & Newport, 2015, 2017).

In a slightly different miniature language learning paradigm, Cook (1988) examined the correlation between the word order in the AP and VP. She trained different groups of participants on VPs that either had consistent subject-object-verb (SOV) order or verb-subject-object (VSO) order and asked participants to extrapolate the unseen word order in the AP. She found that learners showed a general preference for postpositional order in the AP regardless of whether they were exposed to SOV or VSO but learners who were exposed to SOV showed this preference *more strongly* than those who were exposed to VSO. These findings are only partially consistent with the preference for harmonic ordering as participants did not show a harmonic preference in the head-initial (VSO) language. One potential explanation for this mixed pattern of results is methodological shortcomings of the experimental design. Cook (1988) used a translation task where participants were asked to choose an artificial language translation for an English sentence. This might have prompted learners to consciously engage in comparing the word orders between English and the artificial language and to potentially adopt a strategy of choosing an AP word order maximally different from English (i.e., postpositional order). Thus, while there is experimental evidence for a harmonic ordering preference in the nominal domain, the evidence for this preference in the AP and VP word order is less clear-cut.

Here, we extend this line of study and use the miniature language learning paradigm to ask whether individual learners have a harmonic preference in head ordering across the AP and VP. Following Cook (1988), we use the extrapolation paradigm but instead of using a translation task, we teach participants a novel language (by exposing them to visual stimuli accompanied by their descriptions in the novel language) and ask participants to describe scenes in the novel language at test. This design better mimics the scenario of language learning and reduces the potential interference from the learners' native language. Thus, if learners indeed have a preference for harmonic ordering across the AP and VP, we expect our setup to reveal it.

We further examine whether the preference for harmonic ordering is modulated by memory constraints. Harmonic ordering can be considered a type of regularization (i.e., reduction of linguistic variation resulting in a simpler grammar) (c.f., Culbertson & Kirby, 2016). Prior work in the nominal domain has shown that the strength of regularization in adult learners is dependent on the complexity of learning and/or production tasks: Learners are more likely to regularize linguistic variation when the input system is complex (Ferdinand, Kirby & Smith, 2019; Hudson Kam & Newport, 2009) or when lexical retrieval in production is more difficult (Hudson Kam, 2019; Hudson Kam & Chang, 2009). Prior work exploring regularization behavior has typically exposed learners to fairly complex linguistic systems containing 2 to 16 competing forms in the input with no a priori preference for a particular form beyond its frequency in the input. Here we ask whether the strength of harmonic ordering preference is similarly modulated by memory demands (operationalized as lexical retrieval difficulty, following Hudson Kam & Chang, 2009) in a simpler linguistic system with two equally frequent variants in the input, one of which is hypothesized to be preferred as it results in harmonic ordering across the AP and VP and thus in a simpler linguistic system.

2. Experiment 1. In this experiment, we used an extrapolation paradigm to ask whether learners prefer harmonic ordering across the AP and VP. We first extensively trained different groups of participants on miniature artificial languages containing APs that either had prepositional or postpositional word order. We then briefly exposed participants to simple transitive sentences

710

with flexible word order (SOV and VSO orders with equal frequency in the input), asking whether participants would prefer to use the VP word order in their own productions that is harmonic with the AP order they were exposed to (i.e., SOV in the postpositional language and VSO in the prepositional language). We further explored the influence of memory load on learners' harmonic preference (operationalized as lexical retrieval difficulty, see Section 2.3 below). Based on prior work on regularization (Hudson Kam, 2019; Hudson Kam & Chang, 2009), we predicted that the preference for harmonic ordering would become stronger with increased lexical retrieval difficulty.

2.1. PARTICIPANTS. The experiment was administered via FindingFive (FindingFive Team, 2019), an online platform for behavioral research. 78 native speakers of prepositional SVO languages (one native Portuguese speaker, one native Italian speaker, one native Spanish speaker and 75 native English speakers) were recruited to participate via Amazon Mechanic Turk, a crowdsourcing platform. Each participant received $4 for participating in the study, which lasted approximately 40 minutes. Six participants were removed from the analysis for failing to adequately learn the miniature languages (see Section 2.5 below), thus leaving 72 participants for the final analysis.

2.2. MINIATURE INPUT LANGUAGES. Participants were instructed that they would be learning a novel miniature artificial language by watching visual scenes (videos or pictures) accompanied by their descriptions. The languages contained four nouns (CHEF, UMBRELLA, CHAIR, TABLE), four verbs (CHOP, PICK UP, KICK, PUSH), and two locational adpositions (UNDER, BEHIND). All novel words consisted of 1-2 syllables and were phototactically legal in English. These words were synthesized individually using Apple Speech Synthesizer (voice 'Alex') and concatenated into phrases and sentences. To avoid unwanted interferences from specific lexical associations, the assignment of verbs, adpositions, and nouns to the novel labels was rotated across two lists. During training and comprehension tests, each visual scene was accompanied by auditory and text descriptions. During production tests, participants described scenes by typing.

Participants were randomly assigned to learn one of two miniature languages that differed in the type of APs they contained: either a deterministic prepositional language, which contained 100% prepositional phrases (e.g., *barsa sool rizba* corresponding to 'chef behind table' order to describe the scene in Figure 1a) or a deterministic postpositional language, which contained 100% postpositional phrases (e.g., *barsa rizba sool* corresponding to 'chef table behind' order to describe the scene in Figure 1a). Both languages had flexible word order in the VP (50% SOV and 50% VSO). For instance, in both languages, the scene in Figure 1b could be described as *barsa rizba kypse* corresponding to 'chef table push' or *kyse barsa rizba* corresponding to 'push chef table'.

a. AP trial	b. VP trial

Figure 1. Examples of the scenes used in AP (left panel) and VP (right panel) exposure and tests.

2.3. Lexical retrieval difficulty Manipulation. To test whether the preference for harmonic ordering is influenced by memory constraints, we manipulated lexical retrieval difficulty during sentence production by varying whether the novel lexicon was visible or not during production trials, following Hudson Kam & Chang (2009). In the hard lexical retrieval condition, the lexicon was not visible during production trials and learners thus had to retrieve the novel names for each character and action from memory when they described scenes in the artificial language. In the easy lexical retrieval condition, learners were provided the novel lexicon on the screen (in alphabetical order), which they could consult during production trials, thus making lexical retrieval easier (see Figure 2). The lexical retrieval manipulation was implemented only during production trials and did not affect exposure or comprehension trials (training trials were, thus, the same across the two lexical retrieval conditions).

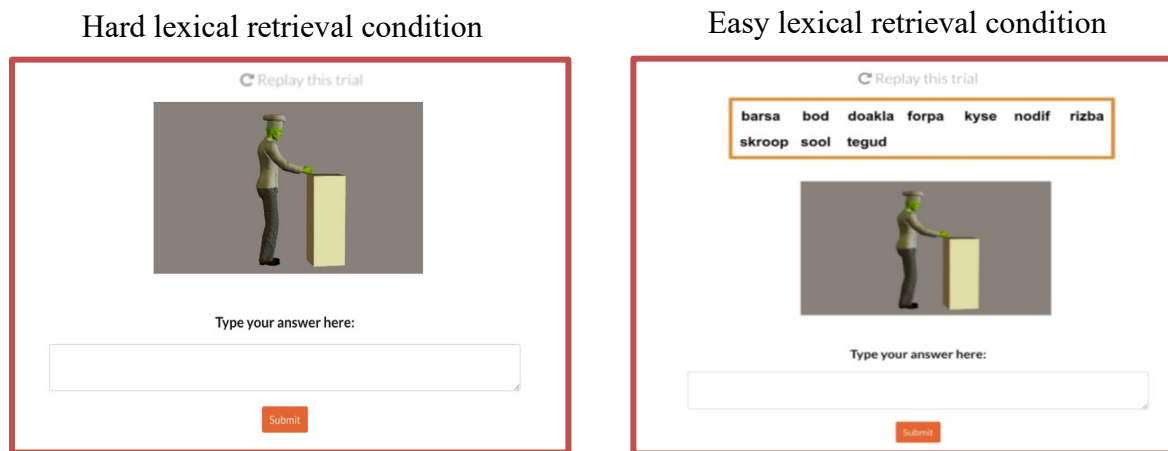Hard lexical retrieval condition   Easy lexical retrieval condition

Figure 2. Sample production trial in Experiment 1. In the hard lexical retrieval condition (left), participants described scenes without lexical prompts. In the easy lexical retrieval condition (right), participants could consult the novel lexicon during production.

When memory load is high (i.e., in the hard lexical retrieval condition) the most accessible form is more likely to be used (Hudson Kam & Chang, 2009), which would correspond to the word order harmonic across the AP and VP in our experiment. Therefore, if memory demands influence regularization in our experiment, we expect learners to show a stronger preference for harmonic ordering when memory load is high (i.e., in the hard lexical retrieval condition) compared to when memory load is low (i.e., in the easy lexical retrieval condition). We further expect learners in the easy lexical retrieval condition to match the input (retain both input word orders with input frequency) as this is a typical preference in adults in the absence of other biases (Hudson Kam and Newport, 2005, 2009; Fedzechkina & Jaeger, 2020).

2.4. procedure. The experiment (eight blocks, each of which focused on different aspects of lexical or grammar learning) had three phases: lexical exposure and test, AP exposure and test, and VP exposure and test.

Lexical exposure and tests. Participants were trained on the novel names for nouns, verbs, and adpositions in isolation. Words belonging to different parts of speech were presented in separate blocks with explicit instructions as to what those words described: names for objects or characters (nouns), actions (verbs), or locations relative to objects (adpositions). The structure of lexical exposure and test blocks was the same for all parts of speech. During exposure, participants viewed pictures or videos of objects, characters, actions, or locations accompanied by the

corresponding labels in the novel language (both orthographic and auditory). Participants then proceeded to the comprehension test, during which they were asked to choose a picture or a video (out of an array of two) that corresponded to the description they heard. Finally, in the production test, participants were asked to provide the labels in the novel language for objects, characters, actions, or locations via typing. Participants received feedback for each lexical comprehension and production trial. Participants were presented with two sets of exposure, comprehension, and production blocks for each part of speech (2-3 times for each lexical item).

AP exposure and test. This phase of the experiment was designed to extensively train participants on either head-initial (prepositional) or head-final (postpositional) APs, depending on the language participants were assigned to learn. Participants were informed that they were going to learn phrases that described where objects/characters are located relative to other objects/characters. During exposure, participants were presented with scenes depicting an object/character under or behind the other object/character (see Figure 1a) accompanied by their oral and orthographic descriptions. Participants then proceeded to the comprehension and production tests following the same procedure as the lexical tests with one exception –participants did not receive feedback on the production test. Participants received one set of exposure (32 trials), comprehension (32 trials), and production (16 trials) blocks.

VP exposure and test. In this phase, participants were briefly (eight trials) trained on simple transitive sentences that did not contain APs. Participants were told they were going to learn sentences that described short videos of a character performing actions on different objects. During exposure, participants viewed simple transitive scenes (50% SOV and 50% VSO word order in the input; same across the prepositional and postpositional languages) accompanied by auditory and orthographic descriptions. Participants then completed a comprehension test (eight trials), during which they were presented with two videos and asked to click on the video corresponding to the sentence they heard. At test (24 trials, the critical block in our experiment), participants were asked to describe simple transitive scenes that did not contain APs via typing (see Figure 1b for an example scene). Participants' word order preferences in the VP on this final test became the dependent measure in our analyses.

2.5. SCORING & EXCLUSIONS. Before turning to our main questions – whether learners prefer harmonic word orders across the AP and VP and whether this preference is modulated by lexical retrieval difficulty, we first describe how the data was scored and what data was excluded from the analysis.

To assess lexical accuracy in production trials, we employed soft-string matching implemented with a Python script to calculate the Levenshtein distance (number of insertions, deletions, and substitutions) from the target word. Content words (nouns and verbs) with a Levenshtein distance higher than two and adpositions with a Levenshtein distance higher than one were considered lexical errors. For instance, the production *scrope* for *skroop* was coded as a lexical error as it involved three edits (one substitution, one insertion, and one deletion) but *skrop* was coded as lexically correct since it only involved one insertion. If AP or VP productions contained additional or missing words from the required target (e.g., *barsa rizba* meaning 'chef table' for *barsa sool rizba* meaning 'chef behind table') or contained two or more lexical errors (thus, preventing us from reliably determining the word order intended by the learner), they were scored as uncodable and removed from the analysis. For all codable productions, we scored the word order used by participants on each trial. All productions containing word orders not allowed by the input were removed from the analysis.

Participants who failed to achieve 70% lexical accuracy on the final production block for each part of speech in the lexical phase (one learner of the prepositional language in the hard lexical retrieval condition and one learner of the prepositional language in the easy lexical retrieval condition) and participants who failed to produce 70% codable productions in AP or VP production blocks (one learner of the prepositional language and one learner of the postpositional language in the hard lexical retrieval condition; two learners of the postpositional language in the easy lexical retrieval condition) were removed from the analysis. This left the data from 72 participants for analysis (hard lexical retrieval condition: 17 participants in the prepositional language, 18 participants in the postpositional language; easy lexical retrieval condition: 19 participants in the prepositional language, 18 participants in the postpositional language).

2.6. RESULTS & DISCUSSION. We now turn to our main questions – whether learners prefer harmonic ordering across the AP and VP and whether this preference is modulated by memory constraints (operationalized as lexical retrieval difficulty). If learners prefer harmonic ordering, then learners exposed to the postpositional language are expected to use more SOV order compared to the learners exposed to the prepositional language. If the strength of this preference is modulated by memory constraints, this preference is expected to be more pronounced when lexical retrieval is hard.

To test these predictions, we employed a mixed logit model (Breslow & Clayton, 2012; Jaeger, 2008) to predict SOV use from language (sum-coded, prepositional = -1, postpositional= 1) and lexical retrieval difficulty (sum-coded; easy = -1, hard = 1) in a full factorial design (all main effects and interactions). The model included the fullest converging random effects structure (random intercepts for participants).

There was no significant main effect of language ($\hat{\beta} = 0.61$, z = 1.71, p =0.09) or lexical retrieval difficulty ($\hat{\beta}= 0.10$, z = 0.27, p = 0.79): Learners used the same amount of SOV in their productions across both languages and lexical retrieval conditions. There was a marginal interaction between language and lexical retrieval difficulty ($\hat{\beta} = 0.66$, z = 1.86, p = 0.07, see Figure 3). Simple effects tests revealed that learners of the postpositional language used significantly more SOV than learners of the prepositional language ($\hat{\beta} = 1.16$, z = 2.56, p < 0.05) when lexical retrieval was hard. By contrast, there was no significant difference between the learners of the two languages in the easy lexical retrieval condition ($\hat{\beta} = -0.05$, z = -0.10, p = 0.92). We found no significant difference in SOV use between the prepositional language learners in the easy lexical retrieval condition and the prepositional language learners in the hard lexical retrieval condition ($\hat{\beta} = -0.55$, z = -1.14, p = 0.25) or between the postpositional learners in the easy lexical retrieval condition and the postpositional learners in the hard lexical retrieval condition ($\hat{\beta} = 0.76$, z = 1.47, p = 0.14).

We further compared participants' SOV use to the input proportion. Wilcoxon signed rank tests revealed that in the hard lexical retrieval condition, learners of the postpositional language used 76% SOV, significantly above the 50% input proportion ($V =111$, p < 0.005), but the prepositional learners matched the input with 48% SOV use ($V=42$, p = 0.83). Learners of both the pre- and postpositional languages in the easy lexical retrieval condition matched the input SOV proportion (prepositional language: 61% SOV, $V = 94.5$, p = 0.17; postpositional language: 57% SOV, $V = 78.5$, p = 0.30).
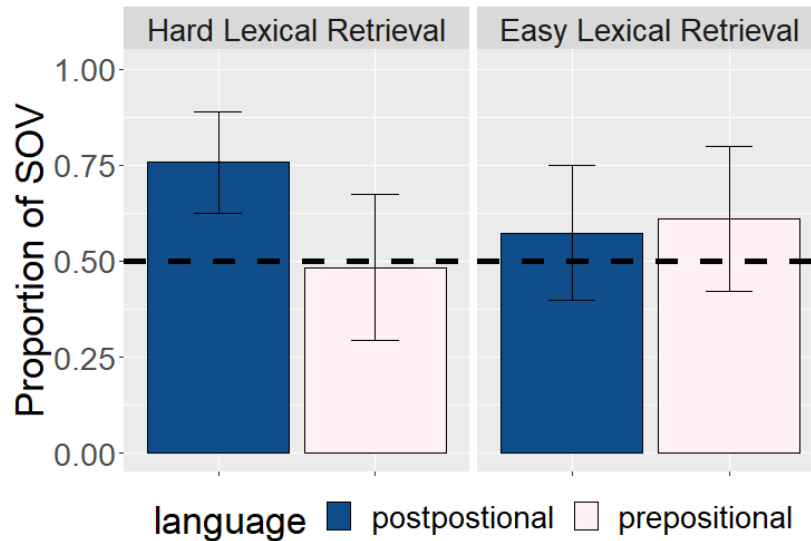
Figure 3. SOV use by language and lexical retrieval difficulty. The error bars show bootstrapped 95% confidence intervals. The dashed line shows the input proportion, same across languages and lexical retrieval conditions.

To summarize, we found that learners of the postpositional language used more SOV compared to the learners of the prepositional language when lexical retrieval was hard but did not differ from the learners of the prepositional language when lexical retrieval was easy, suggesting that memory constraints modulate harmonic ordering preferences in our setup. Our findings, however, provide only partial support for a preference for harmonic ordering in individual learners. While learners of the postpositional language used more SOV compared to prepositional learners in the hard lexical retrieval condition, comparisons to the input distribution indicated that even in the hard lexical retrieval condition, the preference for harmonic ordering was asymmetric across the prepositional and postpositional languages. Specifically, learners exposed to the postpositional language used more SOV in the VP since it is harmonic with postpositional AP order, but learners exposed to the prepositional language did not use more VSO order in the VP despite it being harmonic with the prepositional order in the AP.

One possible explanation for the asymmetry in the harmonic preference across the pre- and postpositional languages is that rather than being sensitive to the position of the head relative to its complements, learners in our experiment are sensitive to the linear order of the head (or non-noun element) across phrases. Consider the AP and VP word orders in the two languages. In the postpositional language, the word order in the AP is head noun--modifier noun--postposition (e.g., 'chef table behind') and the word order in the VP that is harmonic with it is SOV (e.g., 'chef table push'). Thus, in the postpositional language, a strategy of linearly aligning non-noun elements across phrases makes the same predictions as aligning the position of the head relative to its complements. This is not true for the prepositional language, in which the heads of the phrases are not linearly aligned (here, the AP order is head noun--preposition--modifier noun [e.g., 'chef behind table'] and the VP order harmonic with it is VSO [e.g., 'push chef table']). In the prepositional language, the strategy of linearly aligning non-noun elements would not work and learners would be expected to match the input distribution, consistent with learners' preferences in Experiment 1. In Experiment 2, we test whether learners' preferences observed in Experiment 1 can be explained by a preference to linearly align non-noun elements across phrases.

715

3. Experiment 2. In this experiment, we exposed learners in the hard lexical retrieval condition to a prepositional language that had a preposition--modifier noun--head noun word order in the AP. That is, to express the meaning of CHEF BEHIND TABLE, this language used the order 'behind table chef' (as opposed to 'chef behind table' in the prepositional language in Experiment 1). This modification made the non-noun element of the AP (preposition) and the VP (verb) align in the initial position. If the asymmetry in the harmonic preference across the postpositional and prepositional languages observed in Experiment 1 is explained by learners' sensitivity to the linear order of non-nouns across phrases, we expect learners of the prepositional language in Experiment 2 to use significantly less SOV compared to the leaners of the prepositional language in Experiment 1. Apart from this change, all aspects of the input miniature languages, procedure, and scoring were kept the same as in the hard lexical retrieval condition in Experiment 1.

3.1. PARTICIPANTS. 20 self-reported native speakers of a predominantly prepositional SVO language (19 native English speakers and one native Romanian speaker) were recruited from Amazon Mechanic Turk via the FindingFive platform for behavioral research. Each participant received $4 for participating in the experiment, which took approximately 40 minutes. Two participants were excluded from the analysis for failing to adequately learn the language (see Section 2.5 above), which left 18 participants for the analysis.

3.2. RESULTS & DISCUSSION. If learners indeed linearly align the position of non-nouns across phrases, we predict that in Experiment 2, in which the position of the adposition in the AP was linearly aligned with the verb in VSO order, learners should significantly reduce the use of SOV (equivalently increase the use of VSO) compared to the prepositional language in Experiment 1.
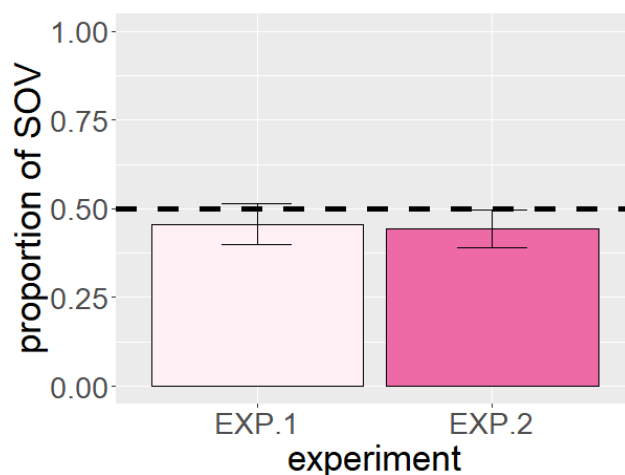

Figure 4. SOV use in the prepositional language (hard lexical retrieval condition) in Experiments 1 and 2. The error bars show bootstrapped 95% confidence intervals. The dashed line shows the input proportion (same across the two experiments).

We compared learning outcomes in Experiment 2 to the learning outcomes in the prepositional language in the hard lexical retrieval condition in Experiment 1. We employed a mixed logit model to predict SOV use in the prepositional language from the experiment (sum-coded, Experiment 1 =1, Experiment 2 = -1). The model included the fullest converging random effects structure (random intercepts for participants and for items defined as verbs). There was no main effect of experiment ($\hat{\beta}$ =0.30, z = 0.61, p = 0.54, see Figure 4), suggesting that the use of SOV order did not differ between the two prepositional languages. Learners' SOV use in Experiment

2 (41%) did not significantly differ from the input proportion (V = 47.5, p = 0.30). Thus, these results suggest that the asymmetry in the harmonic ordering preferences between the postpositional and prepositional languages in Experiment 1 cannot be fully explained by a preference for the linear alignment of non-noun elements.

4. General Discussion. In this study, we asked whether learners have a preference for harmonic word orders across the AP and VP. We tested whether learners exposed to miniature languages would extrapolate from the postpositional order in the AP to SOV in the VP and from the prepositional order in the AP to VSO in the VP as these orders result in the harmonic ordering of heads with respect to their dependents across the two phrase types. We found only partial support for this preference. In line with our hypothesis and typological data, learners exposed to the postpositional language preferentially produced more SOV than learners exposed to the prepositional language. However, this preference was only manifested when lexical retrieval was hard. Furthermore, only learners of the postpositional language produced more SOV compared to the input; learners of the prepositional language produced SOV and VSO equally frequently (matching the input frequency). The absence of the VSO preference in the prepositional language contradicts the preference for harmonic ordering and typological data. Why might there be such an asymmetry in the preference for harmonic ordering between the two languages? One possible explanation, which we ruled out in Experiment 2, is that participants align the linear positions of non-noun elements. Since learners of the prepositional language, in which the order of non-noun elements was linearly aligned across the AP and VP produced the same amount of SOV as the learners of a prepositional language, in which the order of non-noun elements was not linearly aligned across the AP and VP, this possibility is rather unlikely.

An alternative explanation is that learners' preference for harmonic ordering in the prepositional language was obscured by native (L1) influences. All learners in our experiments were native speakers of prepositional SVO languages, who received only brief training on the novel word orders in the VP. Thus, it is possible that learners of the prepositional language extrapolated an L1 harmonic order in their productions that was not in the input (i.e., SVO). Indeed, participants in Experiment 1 occasionally produced SVO word order. However, SVO use did not differ across the pre- and postpositional languages, suggesting that L1 influences cannot fully explain the observed asymmetry in harmonic ordering preferences between the pre- and postpositional languages.

Finally, it is possible that the asymmetric harmonic preference across the postpositional and prepositional languages is driven by the general preference for SOV order found both cross-linguistically (Dryer, 2013) and in experimental work (Goldin-Meadow, So, Özyürek & Mylander, 2008; Tily, Frank & Jaeger, 2011). In the postpositional language, the preference for SOV works in the same direction as the preference for harmonic ordering. In the prepositional language, the preference for SOV works in the opposite direction from the harmonic ordering preference (which favors VSO), which could potentially result in matching the input frequency of the two orders (50%) in this language. Future work on harmonic ordering between the AP and VP could consider exposing participants to OVS and VOS orders instead of SOV and VSO to avoid the potential influence of SOV preference (although we note that such distribution could potentially create other issues as OVS order is extremely dispreferred cross-linguistically).

Interestingly, the asymmetry in the strength of the harmonic preference between the prepositional and postpositional languages observed in our experiment conceptually replicated the findings of Cook (1988). In Cook's experiment, participants showed a preference for postpositional phrases after SOV exposure but did not show a preference for prepositional phrases after

VSO exposure. In our experiment, we found a preference for SOV after exposure to postpositional phrases but no preference for VSO after exposure to prepositional phrases. It is thus possible that the harmonic ordering preference across the VP and AP is weaker than suggested by the typological data (c.f., Dunn et al., 2011). Alternatively, it is possible that we did not choose the right task to probe this preference. One similarity shared by the current study and the work by Cook (1988) is the use of the extrapolation paradigm, in which participants are first exposed to a particular structure and subsequently tested on the extrapolation of the observed structure to an unobserved one or a structure participants were only briefly exposed to. If the harmonic ordering preference between AP and VP is better explained by domain-specific biases (Baker, 2001) or a domain-general bias for simplicity (Culbertson & Kirby, 2016), the extrapolation paradigm is an appropriate design to reveal such a preference. Under these accounts, learners should be able to determine the headedness of the language based on their experience with the AP and extrapolate it to the VP even though the two phrase types never appear in the same sentence. However, the extrapolation paradigm may not reveal the preference for harmonic ordering between AP and VP if it is driven by processing efficiency (Hawkins, 2014) since participants always learn and produce VPs and APs in isolation and never in the same sentence. To address this issue, future work should consider a different paradigm, in which participants do not extrapolate word orders from one phrase type to another but rather produce utterances that include both the VP and AP.

We further found that the preference for harmonic ordering is modulated by lexical retrieval difficulty: Participants showed this preference when lexical retrieval during production was difficult but they probabilistically matched the input when lexical retrieval was easy. This parallels prior work in the nominal domain showing that adult learners regularize the linguistic system when lexical retrieval is difficult but probabilistically match the input otherwise (Hudson Kam, 2019; Hudson & Chang, 2009). Thus, we replicate prior work on regularization in a different paradigm, in which competing variants (i.e., SOV and VSO) are equally frequent in the input (in contrast to previous studies, in which competing variants differed in input frequency), providing further support for the influence of lexical retrieval difficulty on regularization. Our lexical retrieval difficulty manipulation was only implemented for production trials and did not affect training trials, thus suggesting that the preference for harmonic ordering might originate in production. This is consistent with prior work (Hudson Kam, 2019) showing that only increased lexical retrieval difficulty during production but not during learning leads to increased regularization. We leave it to further work to explore the locus of the harmonic ordering preference.

In sum, we used the miniature artificial language learning paradigm to examine whether learners prefer harmonic word orders across the AP and VP and whether this preference is influenced by the difficulty of lexical retrieval in production. Our findings provide only partial support for the harmonic ordering preference. We find that learners of the postpositional language show a harmonic preference (and only when lexical retrieval is hard) but learners of the prepositional language show no such preference regardless of lexical retrieval difficulty. We also find that the preference for harmonic ordering is modulated by the difficulty of lexical retrieval: Learners of both languages probabilistically matched the input distribution once the lexical retrieval difficulty was reduced, consistent with prior studies on regularization in the nominal domain.

References

Baker, Marker C. 2001. *The atoms of language: The mind's hidden rules of grammar*. New York: Basic books.

Breslow, Norman E. & David G. Clayton. 2012. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88(421). 9–25. http://doi.org/10.1080/01621459.1993.10594284.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT press.

Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5). 489–509. https://doi.org/10.1017/S0140525X08004998.

Cook, Vivian J. 1988. Language learners' Extrapolation of Word Order in Micro-Artificial Languages. *Language Learning* 38(4). 497–529. https://doi.org/10.1111/j.1467-1770.1988.tb00165.x.

Culbertson, Jennifer & Simon Kirby. 2016. Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology* 6. 1964. https://doi.org/10.3389/fpsyg.2015.01964.

Culbertson, Jennifer & Elissa L. Newport. 2015. Harmonic biases in child learners: In support of language universals. *Cognition* 139. 71–82. https://doi.org/10.1016/j.cognition.2015.02.007.

Culbertson, Jennifer & Elissa L. Newport. 2017. Innovation of word order harmony across development. *Open Mind* 1(2). 91–100. https://doi.org/10.1162/opmi_a_00010.

Culbertson, Jennifer, Paul Smolensky, & Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122(3). 306–329. https://doi.org/10.1016/j.cognition.2011.10.017.

Cysouw, Michael. 2010. On the probability distribution of typological frequencies. In Christian Ebert, Gerhard Jäger & Jens Michaelis (eds.), *The mathematics of language* (MOL 2009, 2007. Lecture Notes in Computer Science, Volume 6149). Berlin: Springer. https://doi.org/10.1007/978-3-642-14322-9_3.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138. https://doi.org/10.2307/416370.

Dryer, Matthew S. 2013. Order of subject, object and xerb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available at http://wals.info/chapter/81. Accessed on 2020-03-01.)

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Rusell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, 79–82. https://doi.org/10.1038/nature09923

Fedzechkina, Maryia, Becky Chu & T. Florian Jaeger. 2018. Human information processing shapes language change. *Psychological Science* 29(1). 72–82. https://doi.org/10.1177/0956797617728726

Fedzechkina, Maryia, T. Florian Jaeger & Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. Proceedings of the National Academy of Sciences of the United States of America, 109(44), 17897-17902. https://doi.org/10.1073/pnas.1215776109

Fedzechkina, Maryia, Elissa L. Newport & T. Florian Jaeger. 2016. Miniature artificial language learning as a complement to typological data. In In Lourdes Ortega, Andrea E. Tyler, Hae In Park & Mariko Uno (eds.), The usage-based study of language learning and multilingual-ism. 211–232. Washington, DC: Georgetown University Press.

Fedzechkina, Masha, & T. Florian Jaeger. 2020. Production efficiency can cause grammatical change: Learners deviate from the input to better balance efficiency against robust message transmission. *Cognition* 196. 104115. https://doi.org/10.1016/j.cognition.2019.104115.

Ferdinand, Vanessa, Simon Kirby & Kenny Smith. 2019. The cognitive roots of regularization in language. *Cognition* 184. 53–68. https://doi.org/10.1016/j.cognition.2018.12.002.

Ferrer i Cancho, Ramon. 2004. Euclidean distance between syntactically linked words. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 70(5). 056135. https://doi.org/10.1103/PhysRevE.70.056135.

FindingFive Team. 2019. FindingFive: A web platform for creating, running, and managing your studies in one place. FindingFive Corporation (nonprofit), NJ, USA. https://www.findingfive.com.

Finley, Sara & William Badecker. 2012. Learning biases for vowel height harmony. *Journal of Cognitive Science* 13(3). 287–327. https://doi.org/10.17791/jcs.2012.13.3.287.

Fodor, Janet D. 2001. Setting syntactic parameters. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*. 730–767. Oxford: Blackwell. https://doi.org/10.1002/9780470756416.ch23.

Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 105(27). 9163–9168. https://doi.org/10.1073/pnas.0710060105.

Greenberg, Joseph. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), Universals of language. 73–113. Cambridge, MA: MIT Press.

Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199664993.001.0001.

Hudson Kam, Carla L. 2019. Reconsidering retrieval effects on adult regularization of inconsistent variation in language. *Language Learning and Development* 16(1). 43–48. https://doi.org/10.1080/15475441.2019.1634575.

Hudson Kam, Carla L. & Elissa L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2). 151–195. https://doi.org/10.1207/s15473341lld0102_3.

Hudson Kam, Carla L. & Elissa L. Newport.. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59(1). 30–66. https://doi.org/10.1016/j.cogpsych.2009.01.001.

Hudson Kam, Carla L. & Ann Chang. 2009. Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning Memory and Cognition* 35(3). 815–821. https://doi.org/10.1037/a001509 .

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. http://doi.org/10.1016/j.jml.2007.11.007.

Jaeger, T. Florain, Peter Graff, William Croft & Daniel Pontillo. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15(2). 281–319. https://doi.org/10.1515/LITY.2011.021.

Kirby, Simon, Hannah Cornish, & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language.

Proceedings of the National Academy of Sciences of the United States of America (PNAS) 105(31). 10681–10686. https://doi.org/10.1073/pnas.0707835105.

Rafferty, Anna N., Thomas L. Griffiths & Dan Klein. 2014. Analyzing the rate at which languages lose the influence of a common ancestor. *Cognitive Science* 38(7). 1406–1431. https://doi.org/10.1111/cogs.12112.

Tily, Harry J., Michael C. Frank & T. Florian Jaeger. 2011. The learnability of constructed languages reflects typological patterns. *Proceedings of the Annual Meeting of the Cognitive Science Society* 33. 1364–1369.

Travis, Lisa M. 1984. *Parameters and effects of word order variation*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Whitman, John. 2008. The classification of constituent order generalizations and diachronic explanation. In Jeff Good (ed.), *Linguistic universals and language change*. 233–252. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199298495.003.0010.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30(5). 945–982. https://doi.org/10.1207/s15516709cog0000_89.