THE ROLE OF SIMPLICITY IN WORD ORDER HARMONY

DANIELLE NAEGELI [1*], RACHEL KINDELLAN [1*], KENNY SMITH [1], AND JENNIFER CULBERTSON[1]

[*]Joint First Authors, Corresponding Authors: *Danielle.Naegeli@ed.ac.uk*, *Rachel.Kindellan@ed.ac.uk*

[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK

Previous research has demonstrated that the typologically attested preference for word order harmony, consistent order of dependents relative to their head, is reflected in individual learning behaviour in artificial language learning. To address the hypothesis that the word order harmony bias is driven by a more general bias for simplicity, we compared the strength of the word order harmony bias in one- and two-modifier contexts. It was predicted that the harmony bias would be stronger in the one-modifier conditions than in their two-modifier counterparts given that in one-modifier conditions the harmonic grammar is simpler than non-harmonic grammar, whereas in two-modifier phrases they are equally simple. While participants did not exhibit a uniform harmony bias in any condition based on frequency of majority order production, entropy calculations showed a difference in regularisation of harmonic and non-harmonic orders in the one-modifier conditions but not in the two-modifier conditions. The results presented here provide tentative evidence to support the hypothesis that a bias for word order harmony is rooted in simplicity.

1. Introduction

The preference for word order harmony–consistent order of heads and their dependents–has been well-attested cross-linguistically (Greenberg 1963, Dryer 1992). For example, languages with post-nominal adjectives also tend to have post-nominal numerals (Dryer, 2013a: Dryer, 2013b). Recent research has used artificial language experiments to study whether the typological preference for harmony in the noun phrase can be explained by a cognitive bias, e.g., for simplicity (Culbertson et al. 2012, Culbertson & Newport 2015, Culbertson & Kirby 2016); as a result of the transmission of languages through repeated generations of learners, languages then evolve to reflect these biases in learning.

Here we investigate the role of simplicity in driving the preference for harmony by focusing on a potential difference between grammars which generate noun phrases with a single modifier (e.g., an adjective *or* a numeral) and grammars which generate phrases containing multiple modifiers (e.g., an adjective *and* a numeral). Crucially, grammars producing harmonic one-modifier phrases can be shown to be representationally simpler than grammars producing non-harmonic counterparts. However, there is no clear difference in simplicity for harmonic and non-harmonic two-modifier grammars. If the harmony bias is rooted in simplicity, then a preference for word order harmony should be apparent when phrases contain only a single modifier, but weaker (or not present at all) when phrases contain two modifiers.

1.1. Previous experimental findings on noun phrase harmony

Several recent studies have conducted artificial language experiments to investigate whether learners prefer harmonic patterns of noun phrase word order. Culbertson & Newport (2015, 2017) showed that children taught a miniature artificial language comprised of phrases containing either a noun and an adjective *or* a noun and a numeral exhibited a strong preference for harmonic orders (i.e. where the adjective and numeral were placed consistently after the noun). In particular, they were better at learning harmonic input patterns, and shifted non-harmonic patterns to harmonic ones. Culbertson, Smolensky & Legendre (2012) showed that, albeit less strong, adults also have a similar harmony preference in this domain. Saldana, Smith, Kirby & Culbertson (2018) also found a preference for harmonic orders when participants were trained on noun phrases that had either one or two modifiers.

Culbertson & Kirby (2016) argued that the preference for harmonic orders may be rooted in a bias for patterns with simpler (more compressible) grammars. For example, a language which has post-nominal adjectives and post-nominal numerals can be captured by a general rule, e.g., NP → Noun Modifier (where → means 'expands to'). By contrast a non-harmonic language with post-nominal adjectives but pre-nominal numerals requires two specific rules, one for each modifier type, e.g., NP → Noun Adjective; NP → Numeral Noun. The harmonic grammar is shorter than the non-harmonic one, therefore it is simpler. This accords well with previous experimental results, where learners were trained (either exclusively or initially) with phrases containing only a single modifier.

However, there is no straightforward difference in simplicity for grammars producing two-modifier phrases. A learner must learn a rule for placement of the modifiers with respect to the noun as well as with respect to each other. Grammars producing two-modifier phrases following any order therefore consist of one specific rule (e.g., NP → N Adj Num, or NP → Num N Adj). If a cognitive preference for simplicity drives the typological prevalence of harmonic patterns, then this suggests a special role for single modifier phrases (cf. Hawkins 1994). In order to investigate this, we conducted an artificial language learning experiment where learners were exposed to either single modifier phrases or to two-modifier phrases. We predicted that a preference for harmonic orders would be stronger in participants in the one-modifier conditions than those in the two-modifier conditions.

2. Experimental test of the role of simplicity in harmony

2.1. Method

The experiment was closely modelled after Culbertson et al. (2012), however here we used a two-by-two design, manipulating whether the input language was predominantly harmonic or non-harmonic *and* whether participants were trained on one- or two-modifier phrases. The four conditions are illustrated in Table 1. We chose one harmonic pattern (both modifiers post-nominal) and one non-harmonic pattern (post-nominal adjective and pre-nominal numeral) in both one- and two-modifier contexts. As in Culbertson et al. (2012) each condition had a majority order used in the input. For the one-modifier conditions, the majority order for each modifier type was used 70% of the time (a direct replication of Culbertson et al. 2012). For the two-modifier conditions, there is single majority order, used 50% of the time, with all other possible orders used 10% of the time each.[1]

Table 1. Word order frequencies in input languages in each of the four experimental conditions shown by column. The majority word orders for each condition are in bold. In this two-by-two design, there are harmonic and non-harmonic majority orders in one- and two-modifier languages.

One-Modifier Conditions			*Two-Modifier Conditions*		
	Harmonic	*Non-Harmonic*		*Harmonic*	*Non-Harmonic*
Num **N**	0.3	**0.7**	**N** Adj Num	**0.5**	0.1
Adj **N**	0.3	0.3	**N** Num Adj	0.1	0.1
N Num	**0.7**	0.3	Num Adj **N**	0.1	0.1
N Adj	**0.7**	**0.7**	Adj Num **N**	0.1	0.1
			Num **N** Adj	0.1	**0.5**
			Adj **N** Num	0.1	0.1

Stimuli. The lexicon used in this experiment, a subset of that used in Culbertson et al. (2012), is comprised of four nonsense nouns corresponding to novel objects lacking an English label (“blifona”, “grifta”, “nerka”, “wapoga”), three nonsense adjectives (“cherg”,”geej”, “fush”) referring to furry, green and blue, and three

[1] To justify these input frequencies, consider the likelihood that a participant in the one-modifier harmonic condition would see a post-nominal adjective *and* a post-nominal numeral: NMod (0.7) x NMod (0.7) is equal to NModMod (~0.5).

nonsense numerals (“derf”,”kez”,”glawb”) referring to two, three and four. Visual stimuli were identical to those used in Culbertson et al. (2012), extended to two-modifier descriptions as illustrated in Figure 1.

Procedure. Participants were instructed that they would be learning part of a new language. Then they were trained on vocabulary and phrases in the language, as described below. The input for each participant was determined by the condition, explained in detail below.

Phase 1: Vocabulary Training. First participants were trained on nouns in isolation. On each trial a single grayscale noun was presented along with auditory and orthographic labels (40 trials total, 10 for each noun, random order, see Figure 1A). This was followed by a testing phase in which participants were asked to verbally produce a label for a noun presented on the screen (40 trials total, 10 for each noun, random order). Then participants were trained and tested in the same manner for the modifiers in isolation (60 trials total per phase, 10 for each modifier, random order, see Figure 1B, C).

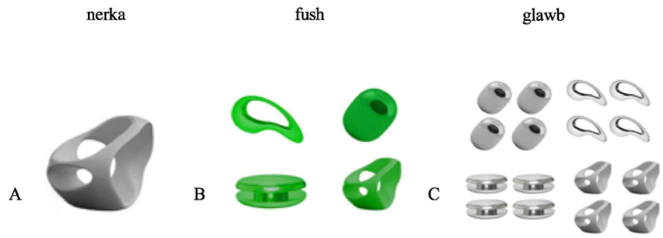


Figure 1: Example vocabulary training images for nouns (A), adjectives (B) and numerals (C). For adjectives, all 4 nouns were presented on screen modified by the relevant property. For numerals set of all 4 nouns were presented on screen in the relevant numerosity.

Phase 2: Noun & Modifier Training. Participants were then trained on phrases in the language. On each trial, a noun modified by either an adjective or a numeral (one-modifier condition) or both an adjective and a numeral (two-modifier condition) was presented on screen and the corresponding phrase was presented auditorily and orthographically. Figure 2 shows an example of each possible trial type.

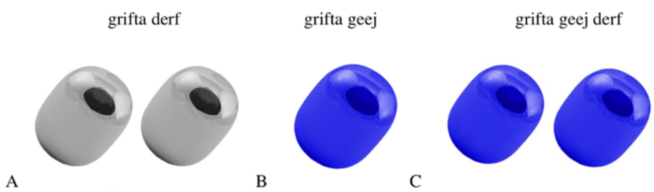
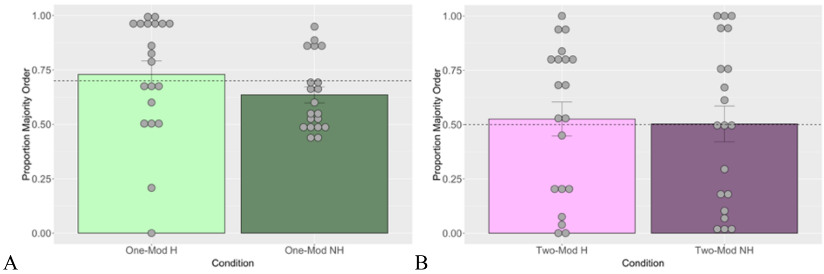


Figure 2: Noun and modifier phrase training examples. For the one-modifier condition numeral (A) and adjective (B) training. For the two-modifier condition numeral plus adjective training (C).

Phase 3 and 4: Picture Matching and Phrase Production. Participants were then tested on their knowledge of phrases in the language. This was done by alternating between comprehension and production. On each comprehension trial a description was presented along with four images: one target image depicting the target noun along with three foils, which varied in one modifier and noun. Participants were then asked to click on the image corresponding to the label (80 trials total, separated into four blocks of 20 trials). On each production trial an image appeared without a label and participants were asked to verbally describe it using the artificial language (80 trials total, separated into four blocks of 20 trials).

3. Results

3.1. Analysis of Majority Order Production



A B

Figure 3. Proportions of majority order use in each condition are represented above. A and B represent one-modifier conditions and two-modifier conditions, respectively. Harmonic (H) and non-harmonic (NH) conditions are depicted on the left and right within the plots, respectively. The dotted lines represent the probability of the majority input order, 0.7 in (A) and 0.5 in (B). Points represent individual points and serve to demonstrate the variability across participants. Standard error is represented by the error bar.

Figure 3 shows how often participants in each condition used the majority order on test trials relative to the input frequency. A binary logistic regression model predicting use of majority order by harmony and number of modifiers (with by-participant random effects) revealed a significant effect of number of modifiers (β = 0.517±0.252, p = 0.040), simply reflecting the difference in majority order in the input, but no significant effect of harmony (β = 0.161±0.252, p = 0.523), and no interaction between harmony and the number of modifiers (β = 0.270±0. 0.252, p = 0.284, respectively). This result is not in line with our predictions as it does not indicate a bias for harmonic word order in the one-modifier condition. However, a closer examination of individual data points reveals that behaviour across conditions was not uniform, and in fact has in some cases bimodal distribution. Figure 4 A and B show the different distributions of outcomes across harmonic and non-harmonic conditions for participants trained on one-modifier phrases; these results are reminiscent of Culbertson et al.'s (2012) results, in that participants in the harmonic condition tend to regularise using the majority order whereas very few participants in the non-harmonic condition regularise using the majority order. Figure 4 C and D show the different distributions of outcomes across harmonic and non-harmonic conditions for participants trained on two-modifier phrases. Like the one-modifier conditions, it appears that regularisation using the majority order is stronger in the harmonic than the non-harmonic order, though the difference in dispersion is less defined than between the one-modifier conditions. To explore these individual-level effects, a second analysis was adopted.

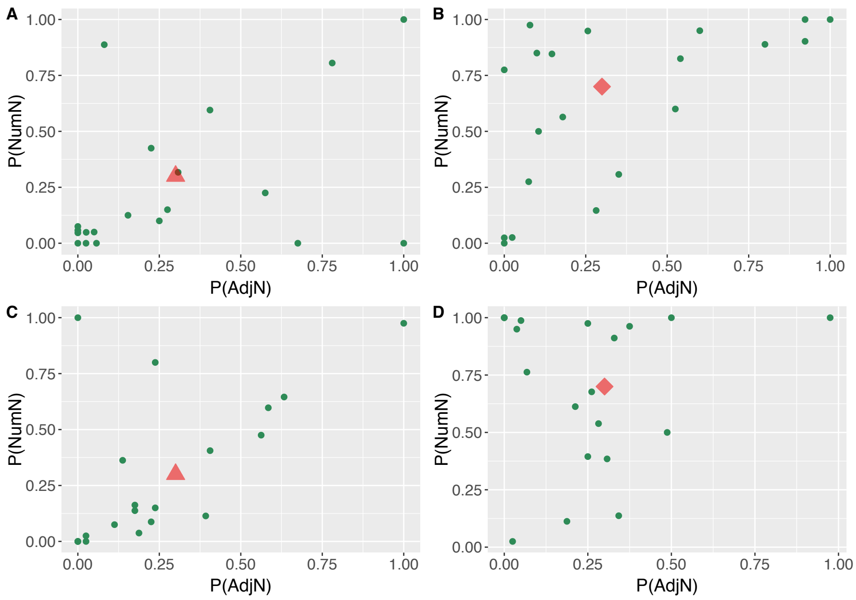


Figure 4. Individual participant outcomes as represented by probability of producing NumN (y-axis) and AdjN (x-axis) in the harmonic (A, C) and non-harmonic (B, D) conditions, for one-modifier (A, B) and two-modifier (C, D) languages. The input word order frequencies are represented by the triangle in (A, C) and by the diamond in (B, D).

3.2. Entropy Analysis

Shannon entropy provides a standardised measure of variation in a system (Shannon, 1948), calculated by taking the sum across all variants in a dataset of the probability of that variant times the log of that probability. We used this measure to assess change in variation in the use of the majority order from the input to the output language. Entropy of the input languages was first calculated. For the one-modifier languages the majority order has probability 0.7 and the minority order 0.3, yielding entropy 0.88. For the two-modifier languages the majority order has probability 0.5 and all minority orders combined have probability 0.5, yielding an entropy of 1. Then, entropy of each participants' output was calculated by collapsing all word orders into either majority order or minority order. This measure allowed for examination of change in variation by use of the majority order. Finally, change in entropy from input to output for each participant was calculated, shown in Figure 5. Unlike the analysis of majority order production reported above, this change in entropy calculation captures the fact that participants in the non-harmonic one-modifier condition maintain variation more and regularise using the majority order less than the one-modifier harmonic counterpart. By contrast, most of the participants in the one-modifier harmonic condition regularise on the majority input order, therefore reducing variation and entropy. This difference between harmonic and non-harmonic orders in the one-modifier conditions is as predicted; it shows a greater reduction in variation by overuse of the harmonic order. Also in line with our hypothesis, no such difference between harmonic and non-harmonic languages is observed in the two-modifier conditions.

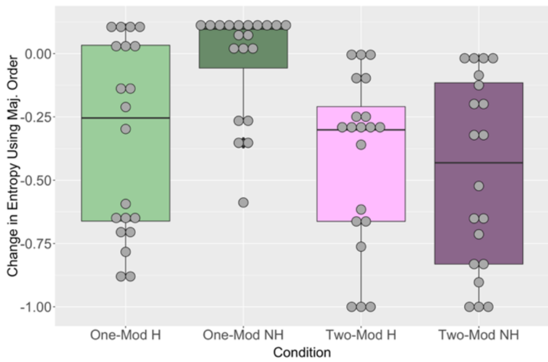


Figure 5. Change in entropy (y-axis) calculated using majority order represented per condition (x-axis). Points represent individual participants. Mean is represented by the bold horizontal line.

Change in entropy across conditions was analysed using a linear regression model. This model confirmed a significant effect of harmony ($\beta = 0.314 \pm 0.106$, $p < 0.01$) and an interaction between harmony and number of modifiers, such that the effect of harmony is stronger in the one modifier conditions than in the two modifier conditions ($\beta = -0.373 \pm 0.149$, $p = 0.015$)[2]. In other words, as predicted, a harmony preference was detected in the one-modifier conditions but not the two modifier conditions.

---

[2] Entropy Change ~ Harmony * Number of Modifiers

4. Discussion

Here we have reported the results of an artificial language learning experiment testing the hypothesis that a cognitive bias for harmony is driven by simplicity. This hypothesis predicts that a bias for harmony should be found (more strongly) across single-modifier phrases, since a grammar encoding harmonic order in such phrases is representationally simpler than one encoding non-harmonic order. By contrast, a grammar with harmonic order in phrases with multiple modifiers is not straightforwardly simpler to represent than one with non-harmonic order. Following previous work (Culbertson et al. 2012), participants were either trained on a predominantly harmonic language or a non-harmonic one. In addition, half of our participants were trained on one-modifier phrases only, and the other half on two-modifier phrases only. An analysis of majority order use failed to reveal the predicted difference between one- and two-modifier conditions in the preference for harmony. However, this analysis is somewhat misleading given that individual participants behaved quite differently from one another (e.g., see also Culbertson & Newport 2015). To deal with this variation, we turned to Shannon entropy to measure the extent to which participants across conditions reduced the variability in word order (comparing the use of majority input order vs other orders). This analysis revealed the predicted effects: significantly more reduction of entropy using majority pattern in the harmonic one-modifier condition compared to the non-harmonic one-modifier condition, but no difference in the two-modifier conditions.

Interestingly, the overall effect of harmony found in our one-modifier conditions appears to be weaker than the effect reported in Culbertson et al. (2012), who found a difference between harmonic and non-harmonic conditions even in the analysis of majority order production. Our guess is that this reflects a difference in the difficulty of acquiring the systems; the size of the lexicons used here was much smaller (4 nouns and six modifiers) compared to Culbertson et al (2012) (10 nouns and 10 modifiers). This may lead to a reduction in the overall effect of simplicity of the grammar, as participants spend fewer cognitive resources on vocabulary learning. Previous studies have also shown that greater cognitive load may lead to stronger regularisation behaviour (Hudson Kam & Newport, 2005; Ferdinand et al., 2019).

To conclude, while analysis of majority order proportions did not reveal a strong effect of harmony, the entropy analysis indicated that the harmonic order is preferred in one-modifier but not two-modifier contexts. This supports the hypothesis that a preference for harmony is rooted in a more general cognitive preference for simplicity, a well-known and wide-spread bias (Chater and Vinyati, 2003; Culbertson and Kirby, 2016). Given that the vast majority of modified noun phrases learners hear will include only a single modifier (Culbertson et al. under review), these phrases therefore have the potential to shape the evolution of noun phrase word order.

References

Chater, N., & Vitanyi, P. (2003). Simplicity: A unifying principle in cognitive science. Trends in Cognitive Science, 7, 19–22.

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

Culbertson, J. & Newport, E. (2017). Innovation of Word Order Harmony Across Development. *OpenMind, 1*(2) 91-100.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306-329.

Culbertson, J., & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in psychology*,*6*, 1964.

Culbertson, Jennifer, Marieke Schouwstra, and Simon Kirby. (2019). From the World to Word Order: The Link Between Conceptual Structure and Language. *PsyArXiv*. doi:10.31234/osf.io/v7be4.

Dryer, M. (1992). The Greenbergian word order correlations. *Language*, 68(1):81–183.

Dryer, M. (2013a). Order of adjective and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved June, 2019 from http://wals.info/chapter/87.

Dryer, M. (2013b). Order of numeral and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved June, 2019 from http://wals.info/chapter/89.

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

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

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

Hudson Kam, C. L., & Chang, A. (2009). Investigating the cause of language regularization in adults: Memory constraints or learning effects? Journal of Experimental Psychology, 35, 815–821.

Saldana, C., Smith, K., Kirby, S. & Culbertson, J. (2018). Is regularisation uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order. Manuscript submitted for publication.