GLOSSA psycholinguistics

Holtz, A., Culbertson, J., & Kirby, S. (2026). Category-specific and system-wide preferences in competition: Evidence from noun phrase harmony. *Glossa Psycholinguistics*, 5(1): 6, pp. 1–29. DOI: https://doi.org/10.5070/G6011.21176

eScholarship
University of California

Category-specific and system-wide preferences in competition: Evidence from noun phrase harmony

Annie Holtz, Centre for Language Evolution, University of Edinburgh, aholtz@ed.ac.uk
Jennifer Culbertson, Centre for Language Evolution, University of Edinburgh, jennifer.culbertson@ed.ac.uk
Simon Kirby, Centre for Language Evolution, University of Edinburgh, simon.kirby@ed.ac.uk

Typological data show a tendency for languages to exhibit harmonic (i.e., consistent) ordering between heads and dependents. Previous experimental work using artificial language learning experiments has shown that learners prefer harmonic patterns. This suggests that the typological trend for harmony may, at least in part, be driven by a cognitive bias. However, it is well-documented that specific categories sometimes contradict this tendency. Here, we investigate one such case in the domain of the noun phrase. While many nominal dependents exhibit harmony, adjectives and genitives do not: adjectives tend to follow the noun and genitives tend to precede. Previous experiments have identified the existence of cognitive biases that keep these dependents split across the head noun in contexts where there is no conventional language system in place. In this study, we use a silent gesture experiment to examine whether the specific biases that apply to these two dependent types compete with a general preference for harmony in an artificial language learning task. Specifically, we examine whether participants' learning behaviour is consistent with a preference not just for harmony, but for a non-harmonic order where adjectives follow and genitives precede the noun. What we find, instead, is that participants' preference for consistent language systems is not modulated by category-specific biases for prenominal genitives and postnominal adjectives. We discuss the implications of this finding for explanations of typological tendencies which link them to cognitive biases.

1. Introduction

Recent research has explored the way that cognitive biases, which favour or disfavour certain linguistic structures, have contributed to the cross-linguistic typological tendency for *harmonic*, or consistent, word order (Christiansen, 2000; Culbertson, 2012; Culbertson & Kirby, 2016; Wang et al., 2023). This work includes experimental, computational and corpus-based studies examining how these structural biases emerge, and how they affect linguistic behaviour across different age groups and structures. Generally, artificial language learning experiments examining the prevalence of harmonic structure have found that people tend to prefer harmonic languages to non-harmonic ones: both adults and children learn, reproduce and regularise (i.e., extend) harmonic patterns rather than non-harmonic ones (Culbertson & Newport, 2017; Culbertson et al., 2012), even if their native languages are non-harmonic (Culbertson & Newport, 2015; Culbertson et al., 2020a; Wang et al., 2023).

Crucially, the bias to harmonise language structure appears to be particularly active during language *learning*, in other words, when participants are trained on a language system and then have to reproduce that language at test (Holtz et al., 2025). For example, studies such as Culbertson et al. (2012) and Culbertson et al. (2020a) use a regularisation design, where the target language that participants learn exhibits random variation, here, in terms of the word order of modifiers and nouns in a noun phrase. Participants in these studies were native speakers of either English, which has harmonic order in the noun phrase where modifiers precede the noun, or Hebrew or French, which feature some degree of non-harmony in the noun phrase – for example, Hebrew has adjectives after, and most numerals before, the noun. In these studies, some participants learn a language that tends toward harmony (i.e., adjectives and numerals mostly on the same side of the noun), and others learn a language that tends toward non-harmony (i.e., adjectives and numerals tend to appear on different sides of the noun). The results show that when the majority word order pattern in these target languages is harmonic, participants are more likely to extend the harmonic order and eliminate the variability present in the training data, independently of their native language's use of harmonic order (Culbertson et al., 2020a). Crucially, for English speakers, the preference for harmony extends to postnominal harmonic orders, suggesting that English-speaking participants do not simply favour their native language system. Participants in all groups, thus, produce languages that exhibit favoured structural features, like harmony, and do not just reproduce patterns based on similarity to their native languages. The idea that the harmony bias is especially active during learning is perhaps unsurprising if we consider the role that harmony plays in providing generalisable rules for a language system. An idealised harmonic language requires encoding of fewer ordering rules than non-harmonic ones (see the example in **Table 1**), and allows for one high-level rule to dictate head-dependent order across phrase types. In order for the benefits of that type of systematic organisation to be relevant, there must be (at least) two different phrase types or dependents in the same context which can be generalised over. And so, learning tasks where multiple different phrases are taught to participants naturally should be more likely to reveal a preference for harmony.

Table 1: Illustrative example of how a harmonic language can be represented with fewer ordering rules than a non-harmonic language.

Harmonic	**Non-harmonic**
XP ⟶ X …	NP ⟶ N…
	VP ⟶ …V
	PP ⟶ P…

In contrast, studies in which participants have to improvise linguistic structures in the absence of an existing language system do not necessarily find a preference for harmonic orders. For example, Culbertson et al. (2020b) asked English-speaking participants to improvise silent gestures to convey simple objects and their properties, e.g., 'these four spotted toothbrushes'. Participants in this study did not simply reproduce English-like prenominal modifiers. Rather, they produced a wide variety of different gesture orders. However, most of these could be considered non-harmonic, equivalent to something like THESE FOUR TOOTHBRUSHES SPOTTED, or THESE TOOTHBRUSHES SPOTTED FOUR, where only a subset of dependent meanings are on the same side of the object meaning gesture.

The difference between the learning experiments, where a harmony bias has been found, and the improvised gesture experiments, where it has not, may lie in the degree to which there is evidence for a system. In the learning experiments, participants only ever had to produce simple noun phrases (i.e., phrases with one dependent at a time), whereas in the improvisation study, participants had to produce complex noun phrases with several dependents. The simple noun phrases in the learning experiments may have highlighted the structural parallel between elements across noun phrases in that task, strengthening participants' tendency to harmonise, whereas the greater complexity of the meanings that participants had to convey in the improvisational study meant that the structural parallels for harmony were not as clear for the complex noun phrases, reducing participants' tendency to produce harmonic gesture orders. In addition, there is evidence that improvisation is generally guided by a different type of cognitive bias, which targets ordering preferences for *specific linguistic items or categories*, rather than guiding systematic organisation across items (as harmony does). This may arise from the fact that in a typical improvisation task, there is no pre-existing system, so there is no possibility for any system-wide biases like harmony to play a role. This stands in contrast to a learning task, in which participants inductively generalise from a set of items to an underlying system. For example, gesture improvisation tasks reveal that different types of verbs favour the use of different basic word orders. Verbs like *imagine*, which denote intensional events, favour SVO order, whereas verbs like *throw*, which denote extensional events, favour SOV order (Schouwstra & de Swart, 2014). Similarly, recent experimental work suggests that adjectives and genitives are also subject to category-specific ordering preferences. When participants have to improvise gestures including a noun and an adjective, they prefer postnominal adjectives (Culbertson et al., 2020b; Holtz et al., 2025; Jaffan et al., 2020). By contrast, when gestures include nouns and genitives, participants prefer prenominal genitives (Holtz et al., 2025). Interestingly, postnominal adjectives and prenominal genitives are also more common in both spoken and signed languages (see **Table 2**).

Table 2: Typological counts of the order of nouns with adjective and noun with genitive, in spoken and signed languages (Coons, 2022; Dryer, 2013a, 2013b).[1]

Adjective	**N (Spoken)**	**N (Signed)**	**Genitive**	**N (Spoken)**	**N (Signed)**
Postnominal	879	16	Postnominal	468	2
Prenominal	373	10	Prenominal	685	22
Other	110	15	Other	96	10

Why might there be distinct ordering preferences for adjectives and genitives? One possibility is that they are grounded in considerations of language processing and semantics. Postnominal adjectives may be preferred because many of the most common adjectives across languages are *relative*, they depend on the noun to be accurately interpreted. For example, the adjective *small* in phrases like *small ant* and *small house* denote two very different sizes, based on the context provided by the noun. Therefore, having access to the noun prior to encountering the adjective allows for more effective incremental processing than a prenominal adjective (Culbertson et al., 2020b; Jaffan et al., 2020; Rubio-Fernandez et al., 2022; Weisleder & Fernald, 2009). Postnominal adjective placement has, thus, been argued to impart an advantage in semantic processing of complex noun phrases (Culbertson et al., 2012). The contrasting prenominal

[1] Due to the low *N*, it is hard to evaluate the reliability of this tendency for signed languages. The data in this table is based on raw frequencies of languages which are classified rather strictly. Recent research shows that word order tends to exist on more of a gradient than these divisions imply (Levshina et al., 2023), and raw counts also lack factors controlling for language family (Dunn et al., 2011; Hartung et al., 2022). When examining word order correlations while controlling for language family in the sample, the results appear mixed, with some studies finding that typological tendencies are highly lineage-specific (Hartung et al., 2022), while others find more evidence for lineage independence (Jäger & Wahle, 2021).

preference for genitives may, instead, reflect the association between possession and animacy. Possessors tend to be highly animate (Rosenbach, 2008; Silverstein, 1986; Yamamoto, 1999), and animate entities, in turn, have been argued to hold a privileged position in the processing and linearisation of linguistic elements (Dahl, 2008), which leads to them being favoured in specific syntactic roles and early linear positions (Hawkins, 1981; McDonald et al., 1993; Tanaka et al., 2011). The impact of animacy on genitive ordering has been explicitly tested in languages that allow for both prenominal and postnominal genitive structures, with results showing that animate possessors predict the use of the prenominal genitive order (Rosenbach, 2005; Strunk, 2004; van Bergen, 2011). Together, these considerations would favour prenominal genitives and postnominal adjectives.

Typological data on harmony between these two dependents shows evidence of both types of biases at work: while the two harmonic orders are both well-attested, the non-harmonic order which has prenominal genitives and postnominal adjectives is just as common (see **Table 3**). This is interesting, because biases which are active during language learning will, in principle, exert a recurring influence on language structure; individual learning biases of this kind have been shown to be amplified by the process of cultural transmission (Kirby et al., 2007; Thompson et al., 2016). By contrast, category-specific biases – which, based on previous experimental work, appear to be active mainly during improvisation, more akin to language creation than language learning – presumably have less opportunity to influence typology. And yet, these category-specific biases are visible in the typology. However, two recent studies have found evidence that some category-specific biases may also influence language learning tasks. For example, Motamedi et al. (2022) found better learning/regularisation of language systems where the basic

Table 3: Typological counts of the order of noun, genitive and adjective, based on spoken language typological data (Dryer, 2013a, 2013b) and signed language typological data (Coons, 2022).[2] Harmonic patterns are common, but so is a particular non-harmonic ordering, with postnominal adjectives and prenominal genitives (bolded).

Order	**N (Spoken)**	**N (Signed)**
N-Adj / N-Gen	342	–
Adj-N / Gen-N	232	8
N-Adj / Gen-N	**342**	**9**
Adj-N / N-Gen	65	–

[2] The signed language data in this table was collated by the first author based on available data from Coons (2022). The included data is limited to those languages where a clear order could be defined for both dependent types.

word order for intensional and extensional events matched the order favoured by the category-specific biases mentioned above (SOV for extensional events and SVO for intensional events). Further, Do et al. (2022) also found better learning/regularisation of an artificial language with postnominal adjectives, compared to one with prenominal adjectives (see also Culbertson et al., 2012). Together with the typological data, these studies suggest that the category-specific biases for prenominal genitives and postnominal adjectives are not necessarily restricted to improvisational contexts, and could potentially guide learning behaviour.

We explore this possibility using a gesture-based artificial language learning experiment. Artificial language learning with gestural stimuli is a powerful experimental tool, as it allows us to observe cognitive biases that influence participant's linguistic behaviour, while limiting the effect of participants' native language (Goldin-Meadow et al., 2008; Motamedi-Mousavi, 2017). Furthermore, it also enables us to manipulate the experimental task to create language contexts that simulate different stages of the evolution of language (Holtz et al., 2025), thereby examining which cognitive biases are active at different stages of language evolution. In our study, participants are trained on a gesture system with variable word order, similar to previous experiments targeting regularisation of variation during learning (Culbertson et al., 2012; Ferdinand et al., 2019; Motamedi et al., 2022). This design allows us to compare the degree to which participants learn different word order patterns that vary based on how consistently they employ the same word order across dependent types.

Importantly, the artificial languages that participants are trained on vary along two variables, namely, how much of the system exhibits natural orders for adjectives and genitives, and how harmonic the system is across these two dependents (i.e., how aligned the distributions of pre- and postnominal word order are across the two dependent types). Following Culbertson et al. (2020a) and Culbertson et al. (2012) we quantify a preference for harmony as better learning and more regularisation of skewed distributions which are already harmonic. In other words, if participants are better able to pick up and extend a harmonic language, despite noisy word order variation, then we will conclude that harmonic patterns are preferred. We also test whether category-specific biases influence learning. If they do, then we would predict a greater tendency to extend word order systems that align with category-specific biases compared to unnatural orders. Finally, following Smith and Wonnacott (2010), we additionally measure alternative ways in which participants may reduce unpredictable variation in their output, by two types of system-wide regularisation that are distinct from harmony per se. First, we look at the overall use of a single word order across the board. The measure we use for this, described in detail below, would capture regularisation behaviour independent of dependent type; for example, participants might be highly regular in the order they produce for adjectives, but maintain input variation for numerals. This would not reflect harmony, but would, nevertheless, reflect an increase in regularity. Second, we look at context-dependent use of word orders, in which the use of order is predictable if you know the dependent type. The measure we use for this, again described in detail below, explicitly captures regularity based on differences in word order use between dependents.

2. Experiment

The experiment had a between-subjects design, based on previous work employing regularisation designs in ALL (Culbertson et al., 2012; Ferdinand et al., 2019; Motamedi et al., 2022).[3] In these studies, the variable input presented during training is compared to the output that participants produce at test. The prediction is typically that participants are more likely to regularise (or extend) the use of preferred variants, compared to dispreferred variants.

All participants in our study were exposed to a majority and a minority gesture order for both adjective and genitive meanings during training. The majority order appeared in 75% of training trials and the minority order in the remaining 25% of training trials. The main manipulation between the four experimental conditions was in whether the majority order was prenominal or postnominal for each dependent type (see **Table 4** for condition majority/minority orders). Two of the conditions were majority harmonic (both dependents tended to be prenominal, or both tended to be postnominal). Two of the conditions were majority non-harmonic (dependents tended to come on opposite sides of the noun). Of the latter, one (natural condition) matched the hypothesised category-specific biases, with adjectives tending to be postnominal and genitives tending to be prenominal, and other (unnatural condition) did not.

Table 4: Percentage of prenominal and postnominal gesture orders in input per dependent type across conditions.

Condition	**Adjective**		**Genitive**	
Order	**Pre**	**Post**	**Pre**	**Post**
Natural (non-harmonic)	25%	75%	75%	25%
Unnatural (non-harmonic)	75%	25%	25%	75%
Prenominal (harmonic)	75%	25%	75%	25%
Postnominal (harmonic)	25%	75%	25%	75%

[3] The preregistration for this study can be viewed here. Note that the review process and feedback from colleagues during talks has resulted in some changes between the planned analysis and the one presented here.

2.1 Methods

2.1.1 Materials

The experiment was built in the JavaScript library jsPsych (de Leeuw et al., 2023) and ran in participants' web browsers. The experiment had two main sets of stimuli. First, a collection of grayscale digital drawings depicting instances of item ownership (genitives) or items with different patterns (adjectives). The total set comprised 16 images, with every combination of the four items (*book*, *cup*, *hat*, and *scarf*), combined separately with the two possessors (*vampire* and *cyclops*), and the two patterns (*spots* and *stripes*). Each target, therefore, comprised two meaning elements, either a head noun and a possessor or a head noun and a pattern. Examples of the images can be seen in **Figure 1**.

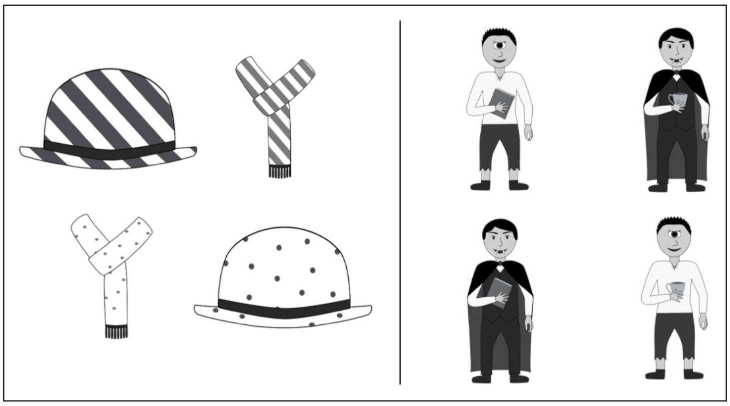

Figure 1: Sample stimuli from an adjective trial (left) and a genitive trial (right).

Paired with each image there were two corresponding gesture videos. The videos showed a model gesturer producing an expression signifying the meanings of the images using a sequence of two gestures (see **Figure 2** for example stills from one of the gesture videos). One gesture represented the head noun and one the adjective/genitive dependent. Each pair of videos differed

(a) Gesture for noun *book*. **(b)** Gesture for adjective *striped*.

Figure 2: Stills of gestures used to denote the meaning *striped book*.

only in the order of these two gestures – in one video, the sequence was head-initial and in the other, it was head-final. Each meaning component was denoted using a two-handed gesture and each sequence ended simultaneously in a neutral position with both hands clasped just below waist-height. All videos were 4.389 seconds long.

2.1.2 Procedure

Participants were randomly assigned to one of the four conditions seen in **Table 4**, and a pseudo-randomised stimuli set consisting of four meanings.[4] The complete stimuli set consisted of two nouns, one randomly selected from the set of "worn" items (*hat* and *scarf*) and one randomly selected from the set of "held" items (*cup* and *book*). These both appeared with one of the genitives and one of the adjectives, respectively, making a stimuli set where two targets were genitive meanings and two were adjective meanings. For example, a participant may have had a stimuli set consisting of the meanings *vampire's book*, *cyclops' hat*, *striped book*, and *spotted hat*. At the start of the experiment, participants were told that they would learn to express ownership and describe items in a made-up sign language. They were then familiarised with the stimuli by being shown two trials including 2 × 2 grids of adjective and genitive meanings, like those in **Figure 1**.

During the training phase, participants saw these same grids, but with one image highlighted by a red square. Under the images, a gesture video associated with the highlighted meaning was played. Each of the four images in the assigned stimuli set acted as targets 8 times each, making 32 training trials. The proportion of prenominal versus postnominal gesture orders in the training videos varied depending on the condition (see **Table 4**), but 6/8 times the gesture video was the majority order for that dependent type in the specific condition, and twice it was the minority order for that dependent type in the specific condition. The training phase progressed automatically, and participants were instructed to sit back and watch the videos.

In the testing phase, participants saw the same image grids as during training, with the target image highlighted as during training. The target meanings in the testing phase were the same as in training. The main difference between the training and testing phase was that, in the testing phase, both possible gesture videos (prenominal and postnominal gesture order) were looping in aligned positions beneath the images (see **Figure 3**). Participants were instructed to "click on the corresponding gesture video" that was like the ones they had seen during training. Between each of the 32 test trials (8 per meaning), participants also clicked a centred "next" button to reset the mouse. Following the testing phase, participants provided English translations for all four target meanings, and gave details regarding their proficiency in signed and spoken languages.[5]

[4] This stimuli set was comparable in size to previous studies using similar methodology, e.g., Motamedi et al. (2022) where Experiment 2a had two meanings and Experiment 2b had six meanings in the training set.

[5] Details regarding spoken- and sign-language proficiency for participants can be found in the OSF repository linked in the data accessibility statement.

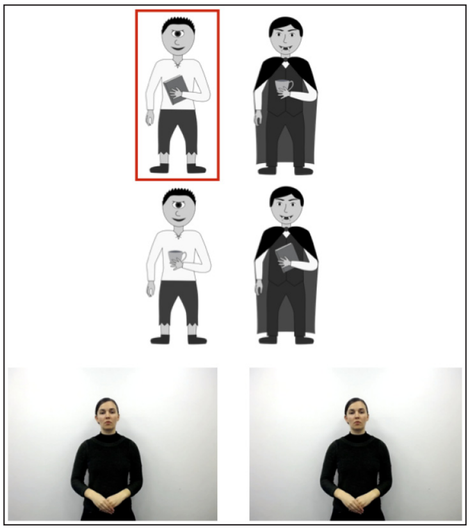

Figure 3: Sample test trial for the genitive meaning *cyclops' book*. The two videos included the prenominal and postnominal gesture order for this meaning.

2.1.3 Participants

211 participants were recruited through the online crowdsourcing platform Prolific. We prescreened participants for the following criteria: 95% previous task approval rate, reported English as their first language, and no previous participation in our experiments/pilots. Participants were paid the equivalent of £8.91 per hour.[6]

2.2 Results

The main predictions for this experiment were as follows. We predicted that participants would generally reproduce the orders they saw in training (i.e., choose the majority gesture order from the training data). Based on previous studies (e.g., Culbertson et al., 2012, 2020a), we predicted better learning in the harmonic conditions (majority prenominal and majority postnominal). As mentioned above, if participants prefer harmonic patterns, they might be more successful at reproducing harmonic input patterns in the presence of noisy variation, *and* they might be more likely to extend, i.e., regularise these patterns. We also predicted that the naturalness of the training orders would modulate learning, such that natural orders would be learned better than unnatural orders. Which of the predictions we set out here is borne out will inform our understanding of when and how system-wide biases like harmony interact with, and are modulated by, category-specific biases. If we observe a better learning of natural orders, but

[6] Following preregistered exclusion criteria, we excluded 8 participants who reported proficiency in a signed language, as well as 3 participants who always pressed the same button during testing. The data in the analysis comes from 50 participants in the natural condition, 53 in the unnatural condition, 46 in the majority prenominal condition, and 51 in the majority postnominal condition.

no tendency for better learning in the harmonic conditions, this would suggest that this task did not activate the system-wide harmony bias, but did tap into category-specific preferences. On the other hand, if we observe a general preference for learning the harmonic systems which is not modulated by naturalness, then this would suggest that category-specific biases do not influence behaviour in this type of task. Of course, if neither prediction is borne out, i.e., there is no difference between conditions in terms of learning, and no effect of naturalness, this would suggest that neither type of bias was active in the task.

2.2.1 Learning

Our first analysis examines differences in how well participants across conditions learned the systems they were trained on. This analysis examines our prediction that harmonic systems are learned better, and can reveal a preference for natural orders in learning if participants also produce more majority orders for a given meaning when the training majority order aligns with the hypothesised preferred order for each dependent type (i.e., prenominal for genitives and postnominal for adjectives). Our dependent measure in this analysis is participants' choice of the majority order from their training data (for each dependent type). **Figure 4** shows the proportion of majority orders selected during the test phase by participants in each condition, as well as the overall grand mean. To analyse this behaviour, we used the brms package (Bürkner, 2017) in R (R Core Team, 2013) to fit a Bayesian linear regression model with a Bernoulli likelihood to participants' production data.[7] The model estimates the log-odds of choosing the majority order (1 indicating a match between majority training order for a given dependent type and the order selected at test) as a function of condition, dependent type, and their interaction. Both fixed effects were deviation coded. Dependent type had two levels (genitive: 0.5, adjective: –0.5), meaning that positive values for the dependent type coefficient would mean higher likelihood of choosing majority order for genitives than adjectives, whereas a negative effect would mean the opposite. The four levels of the condition variable were deviation coded using the contr.sum() function in R, resulting in the coding scheme observed in **Table 5**. Note that when model outputs are reported for the natural condition, i.e., the fourth level in the model, these are calculated by taking the negative sum of the other three levels' coefficients. The same applies for calculating the interactions, where we calculate the fourth level's interaction coefficient by taking the negative sum of the interaction coefficients for the other three levels. The random effect structure included by-participant adjustments to the intercept and to the slopes of dependent type, as well as by-item adjustments to the intercept for dependent type, condition, and their interaction. We included weakly regularising priors (intercept: normal(0, 1.5), fixed effects: normal(0,10) and random effects: normal(0,5)). The model converged, as indicated by all Rhats = 1.00.

[7] The models we preregistered for this study were run using lme4; due to convergence issues and reviewer feedback, we moved to brms, which allowed us to include the maximum random effect structure for our data.

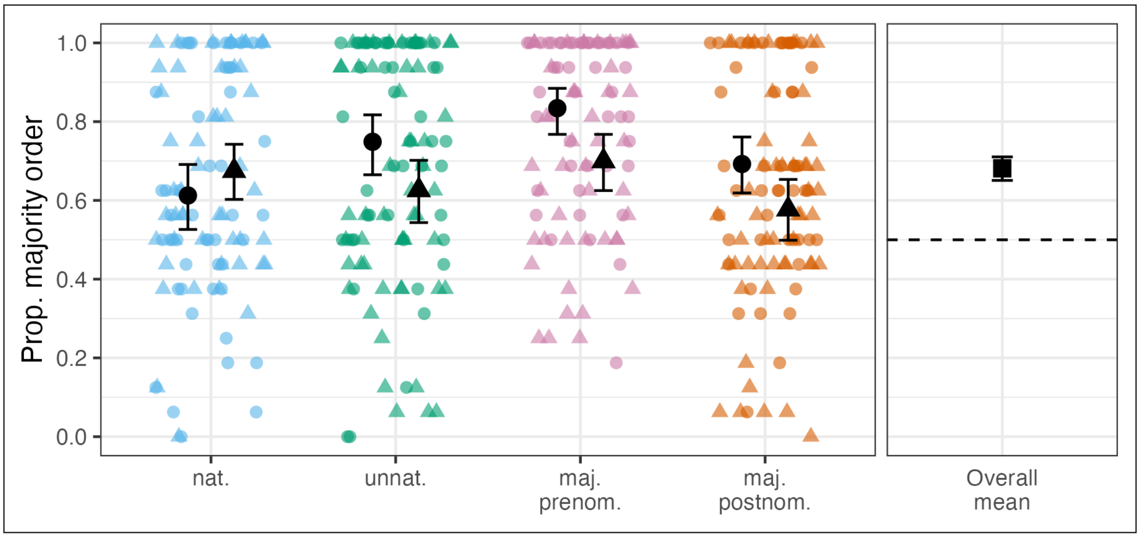

Figure 4: Proportion of test trials where participants selected the majority input order for each condition and dependent type. Right-hand facet shows grand mean. Participants tended to reproduce the majority orders from their training, and this was generally not modulated by naturalness. Error bars show bootstrapped 95% CIs.

Table 5: Contrast coding scheme for the condition variable.

	[1]	**[2]**	**[3]**
unnat.	1	0	0
maj. postnom.	0	1	0
maj. prenom.	0	0	1
nat.	–1	–1	–1

A preference for harmonic systems based on this model structure would thus reveal itself in better learning in the two majority order conditions (main effects in the model). A preference for naturalness, on the other hand, would be supported by a positive main effect for the natural condition, a negative main effect for the unnatural condition, and in the interactions between condition and dependent type in the majority order conditions. A positive interaction for the majority prenominal condition and dependent type would signify better learning of genitives, and a negative interaction for the majority postnominal condition and dependent type would signify better learning of adjectives.

The posteriors from the model are visualised in **Figure 5**, and a summary of model outputs is given in **Table 6**. The mean posterior for model's intercept was 1.37 and has 95% Credible Intervals spanning 1.06 to 1.70, which indicates that, with 95% certainty, participants generally selected the majority order at a rate higher than chance. In other words, they generally reproduced the skewed patterns they were trained on. Of the individual conditions, the only one for which the model placed most of the posterior probability on a value above the grand mean was the majority prenominal condition (mean posterior: 0.67 and CrIs 0.09 to 1.27). The model was unsure of the direction of the effect of learning in the natural condition (mean posterior: –0.35 and CrIs –0.90 to 0.18), and the same was true for the unnatural and majority postnominal conditions. The model places more of the posterior probability mass on a value below the grand mean for the effect of dependent type (mean posterior: –0.77 and CrIs –1.31 to –0.24), reflecting participants' tendency to select fewer majority orders for genitives, compared to the grand mean. The only interaction for which the model placed most of the posterior probability on a value above the grand mean was that between the natural condition and dependent type, although the lower bound of the CrIs was close to 0 (mean posterior: 0.95 and CrIs 0.05 to 1.88), reflecting some uncertainty in the model's estimate that participants' chose *more* majority orders for genitives in that specific condition. Overall, these results indicate that participants learned the systems they were trained on. There was no clear effect of naturalness, since learning was not reliably better or worse in the natural condition and the unnatural condition, respectively, and no reliable interactions were found between condition and dependent type for the majority order conditions. Furthermore, learning was only reliably better in one of the two harmonic conditions, the prenominal one. Because this system aligns with participants' native language order, this finding is, therefore, not strong evidence for a harmony bias.

Table 6: Posterior means and 95% CrIs for the fixed effects of the learning model, given in log-odds space.

	Posterior mean	**95% CrI (lower)**	**95% CrI (upper)**
Intercept	1.37	1.06	1.70
nat.	–0.35	–0.90	0.18
unnat.	0.05	–0.43	0.52
maj. prenom.	0.67	0.09	1.27
maj. postnom.	–0.36	–0.88	0.17
dependent type	–0.77	–1.31	–0.24
nat:dependent type	0.95	0.05	1.88
unnat:dependent type	–0.27	–1.01	0.51
maj. prenom:dependent type	–0.49	–1.49	0.50
maj. postnom:dependent type	–0.25	–1.16	0.62

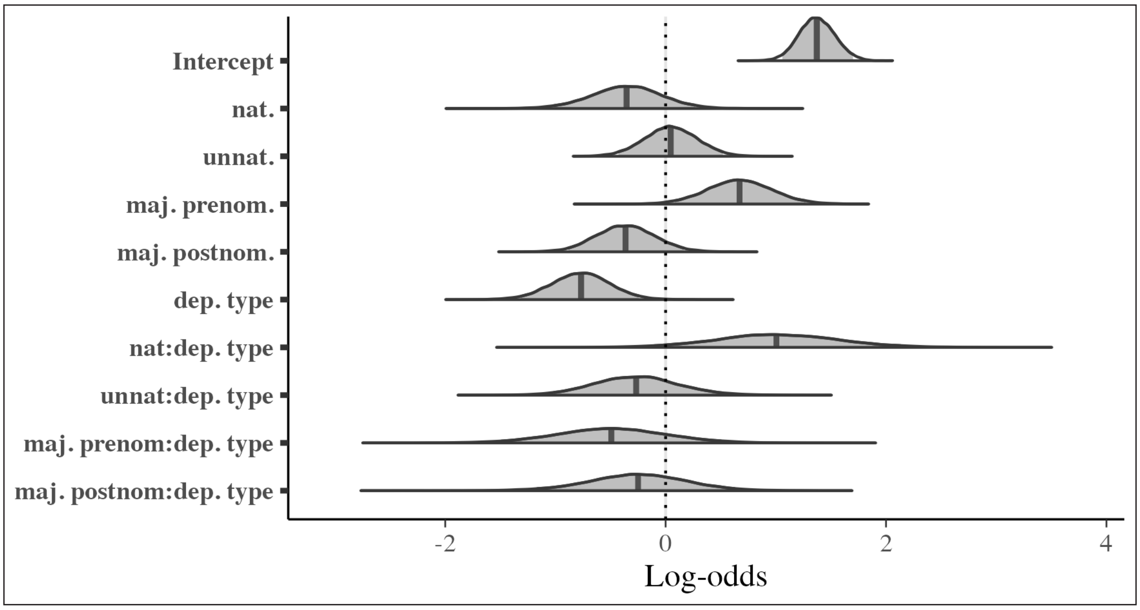

Figure 5: Posterior distributions for each fixed effect and interaction in the Bayesian learning model. Shaded regions show 95% Credible Intervals.

2.2.2 Use of consistent word order

In Section 1, we defined harmony in terms of the alignment of word order across different dependents. Under this definition, the input systems in the majority prenominal and majority postnominal conditions are both perfectly harmonic: the probability of a given order for one dependent type is the same as the probability of that order for the other dependent type, i.e., 75%.[8] Above, we looked for evidence of better learning in these two conditions, and failed to find it. However, another possibility is that participants could extend, or regularise, a single word order across the system. This does not necessarily result in a harmonic system by our definition, as it ignores differences in the precise frequencies with which the word orders are used across dependent types. Nevertheless, it does mean that one order becomes more common across the full system, and reduces the level of uncertainty about what orders are used across said system. We can quantify this behaviour simply as an overall reduction in entropy, following Ferdinand et al. (2019) and Motamedi et al. (2021). Entropy is a measure within Information Theory which is capable of quantifying levels of uncertainty within a system (Shannon, 1948). The entropy (H) of a system is defined as:

[8] This might, at first, glance seem counter-intuitive, since it might seem that the only "perfectly harmonic" system should be one in which the particular order that dependents take is also consistent, e.g., 100% prenominal for both dependent types. However, under the hypothesis that harmony is driven by a bias for simplicity (e.g., Culbertson & Kirby, 2016; Culbertson et al., 2020a) this characterisation makes sense: a system is harmonic if there is no need to posit modifier-specific rules. That is equally true of any system in which the distribution of orders is the same across dependent types. In information-theoretic terms, there is no additional information gained by specifying dependent type.

$$H(V) = -\sum_{v_i \varepsilon V} p(v_i) log_2 p(v_i)$$

where (V) refers to the two gesture variants (prenominal and postnominal orders). All entropy values used in our calculations are based on relative frequency estimation. Both the natural and unnatural training data had an entropy value of 1, as participants were exposed to a 50/50 split of both variants (orders); this means that the non-harmonic conditions are maximally entropic, and entropy cannot increase in these two conditions. The remaining two conditions have a training entropy of 0.8112781. We measured the change in entropy between these input values and the orders that participants selected to see whether participants tended to reduce variation by using more of one gesture order across the system than in the training data. Furthermore, we were interested to see if changes in entropy would be modulated by category-specific biases, such that extension of a given gesture order would be lower in the natural condition compared to the other conditions (indicating influence of category-specific biases acting to preserve the use of different word orders for individual dependent types). **Figure 6** shows the mean change in entropy for each condition as well as the overall mean change.

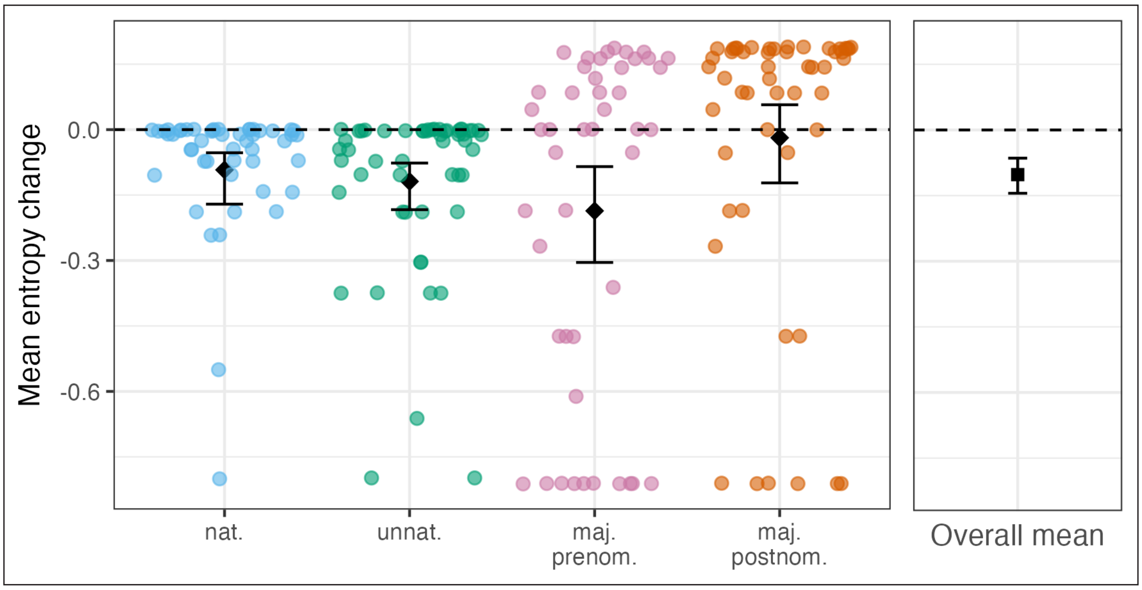

Figure 6: Mean change in entropy between input and output, showing participants' tendency to extend the use of one gesture order, except in the postnominal condition. Right-hand facet shows grand mean change in entropy. Error bars show bootstrapped 95% CIs.

To evaluate whether these changes are reliably greater than zero, we calculated bootstrapped confidence intervals around the reported mean entropy changes for each condition.[9] These were generated using the 'boot' package in R and based on 10,000 samples. Our findings suggest that

[9] Linear models are not reported for this data, as the distribution of entropy in our study is necessarily non-normal.

all but one of the conditions – majority postnominal – show a reliable reduction in entropy. However, since entropy estimations based on sparse data is generally downward biased (and recall that the non-harmonic conditions can only ever become less entropic), we verified these results using Monte Carlo simulations, where the probability of sampling prenominal and postnominal order was set to the training proportions for each condition. This allowed us to compare whether the values for change in entropy that we obtained in the study are reliably different from the general downward trend of the entropy data.[10] Based on 10,000 iterations of the Monte Carlo simulations, mean changes in entropy were extracted across all conditions, as well as individual means for each condition. Z-scores were calculated based on simulated and experimental means, with results showing that change in entropy was reliable overall, as well as for each condition except the majority postnominal condition (see **Table 7** for z-scores). Furthermore, confidence intervals around differences between conditions reveal that the majority prenominal condition is reliably different from all other conditions (nat – majPre: $\bar{x}_a - \bar{x}_b = 0.28$, lower CI = 0.16, upper CI = 0.41; unnat – majPre: $\bar{x}_a - \bar{x}_b = 0.26$, lower CI = 0.14, upper CI = 0.38; majPre – majPost: $\bar{x}_a - \bar{x}_b = -0.17$, lower CI = –0.31, upper CI = –0.03). Similarly, the majority postnominal condition is also reliably different from the natural condition ($\bar{x}_a - \bar{x}_b = 0.12$, lower CI = 0.02, upper CI = 0.23).

Table 7: Experimental means, simulated means and resulting z-scores for change in overall entropy. Z-scores show that experimental means are reliably different from the simulated means (indicating that entropy dropped significantly) in all conditions except the majority postnominal condition.

Condition	Exp. mean	Sim. mean	z-score
Overall	–0.102	–0.020	–12.691
nat	–0.092	–0.017	–22.043
unnat	–0.119	–0.017	–29.555
majPre	–0.186	–0.023	–9.175
majPost	–0.019	–0.023	0.249

[10] The Monte Carlo simulation samples prenominal or postnominal orders based the on proportions of those orders in each input condition, so 50/50 for the natural and unnatural conditions and 75/25 probability in favour of the majority order in the prenominal and postnominal conditions. This means that, for each draw, the probability of choosing the majority order in the harmonic conditions is higher, causing a stronger negative mean in the simulated data. The negative change in the harmonic conditions must, therefore, be greater than this simulated downward bias to be seen as significant, to account for the fact that more negative values can be captured by probabilistic sampling.

2.2.3 Exploratory analysis: Conditional entropy and mutual information

It is important to note that the overall entropy measure above does not differentiate between cases where entropy reduces due to more use of the same word order across *both* dependent types, and cases where that reduction is caused by increased use of a gesture order within the context of one specific dependent type, but not the other. In the second situation, participants are reducing the amount of unpredictable variation between their input and their output by extending or introducing a conditioning context which determines the use of a given order – for example, they always use prenominal order for genitives and always postnominal order for adjectives. As such, a language that employs a *different* order for each dependent, but does this completely regularly, would appear as having maximum entropy by the overall entropy measure, but lacking any unpredictable variation due to the conditioning context. This type of conditioning strategy has been observed in other studies and constitutes an expression of a general tendency of language learners to reduce unpredictable variation (Smith & Wonnacott, 2010; Smith et al., 2017). To examine whether either dependent type or noun was used as a conditioning context by participants, we first calculated two types of conditional entropy for each participant based on this formula:

$$H\left(V|C\right) = -\sum_{c_j \varepsilon C} p\left(c_j\right) \sum_{v_i \varepsilon V} p\left(v_i|c_j\right) log_2 p\left(v_i|c_j\right)$$

The context (C) here stands for dependent type in one calculation, and noun in the other.[11] These calculations of conditional entropy also capture reduction in variability due to harmony (becoming more consistent across the whole system will also result in higher consistency within sub-contexts). To disentangle more consistent use of one gesture order across the full system (analysed in Section 2.2.2) from strategies that employ conditioning contexts, we used the extracted conditional entropy values to calculate *Mutual Information* of selected gesture order for both dependent type

$$Mutual\ information\ (MI) = overall\ entropy – dependent\ type\ conditional\ entropy$$

and head noun

$$Mutual\ information\ (MI) = overall\ entropy – head\ noun\ conditional\ entropy$$

A system with an MI value of 1 would mean that variants (gesture orders) were perfectly conditioned on either the two dependent types or the noun, whereas an MI of 0 indicates that the variability within each context mirrors the whole system (i.e., the system is harmonic). **Figure 7**

[11] The preregistration for this study did not include head noun as a context for conditional entropy, it was included here after noting that lexical contexts is a common strategy for conditioning unpredictable variation (Smith & Wonnacott, 2010).

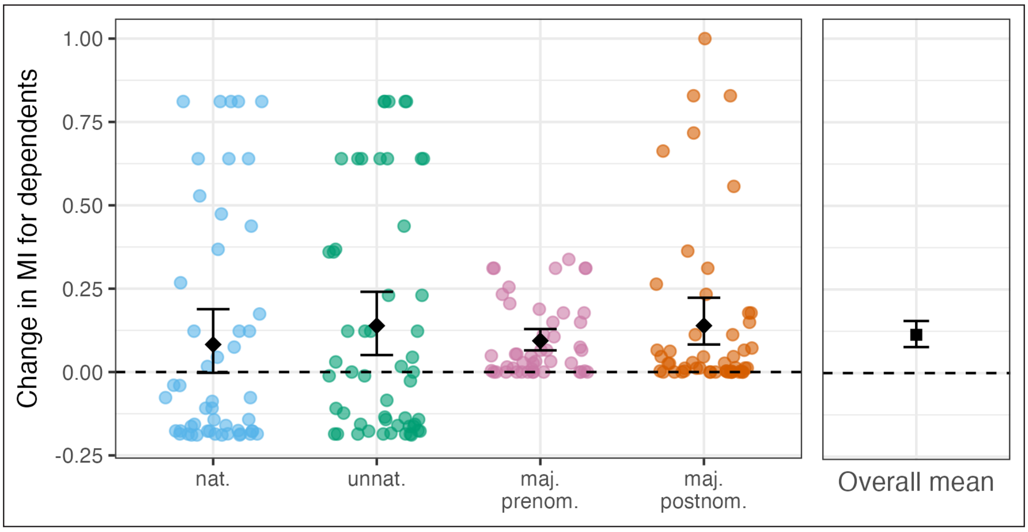

Figure 7: Mean change in MI, based on dependent type, showing reliable change in MI across all conditions except the natural condition. Note that the significance of this measure differs between the presented in this figure and the z-scores comparing simulated and experimental means. Right-hand facet shows grand mean change in MI. Error bars show bootstrapped 95% CIs.

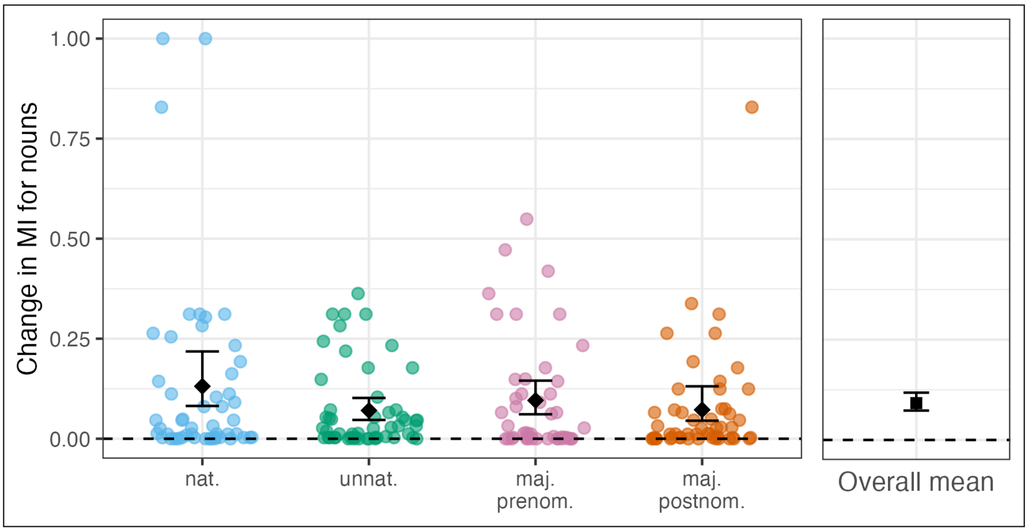

Figure 8: Mean change in MI, based on head nouns, showing slight increase in MI across all conditions. Right-hand facet shows grand mean change in MI. Error bars show bootstrapped 95% CIs.

shows the overall mean MI change (i.e., output MI – input MI) across conditions for dependent type, as well as individual condition means. **Figure 8** shows the overall mean MI change across conditions for head noun, as well as individual condition means. Confidence intervals were generated in the same way as for our overall entropy measure. As with the overall entropy data, we used data from Monte Carlo simulations to verify the reliability of the increase in MI, based on both dependent type and head noun. Based on the simulations, means for both types of conditional entropy were extracted and z-scores calculated. **Table 8** shows z-scores based on dependent type, and **Table 9** shows z-scores based on head noun. Increase in MI is significant across all conditions for both measures. Since an MI of 0 would represent a perfectly harmonic system, if we had observed a reduction in MI in the natural and unnatural conditions, this would have served as additional evidence for a preference for harmony. However, given that we see an increase in MI in these conditions, this suggests that the systems that participants produce in these conditions are not more harmonic than their input system, but they are more regular, i.e., participants do reduce unpredictable variation.

Table 8: Experimental means, simulated means and resulting Z-scores for change in MI, based on dependent type. Z-scores show that experimental means are reliably different from the simulated means (indicating that MI increased significantly) in all conditions.

Condition	Exp. mean	Sim. mean	z-score
Overall	0.114	0.028	12.805
nat	0.083	0.031	2.800
unnat	0.139	0.031	5.834
majPre	0.094	0.025	13.835
majPost	0.139	0.025	22.510

Table 9: Experimental means, simulated means and resulting z-scores for change in MI, based on head noun. Z-scores show that experimental means are reliably different from the simulated means (indicating that MI increased significantly) in all conditions.

Condition	Exp. mean	Sim. mean	z-score
Overall	0.092	0.021	32.639
nat	0.131	0.018	31.946
unnat	0.071	0.018	14.949
majPre	0.096	0.025	14.129
majPost	0.072	0.025	9.436

3. Discussion

This experiment examined whether the previously observed preference for word order harmony during learning would be disrupted by the presence of category-specific ordering preferences. Participants were taught a system with variable word order for nouns with adjectives and genitives which either tended toward harmony (e.g., Noun-Genitive, Noun-Adjective), or tended toward non-harmony (e.g., Noun-Genitive, Adjective-Noun). Crucially, one of the non-harmonic systems aligns with category-specific biases for these dependent types previously found in improvisation tasks (Genitive-Noun, Noun-Adjective). Across our analyses, we found no clear evidence of a preference for harmonic systems. Most clearly, in our analysis of participants' choice of the majority input order across conditions, only the majority prenominal condition exhibited better learning than the other conditions. The majority orders in the prenominal condition align best with the structures that English speakers would use to express descriptive and possessive meanings in their native language (prenominal adjectives and prenominal genitives, respectively). Better learning in this condition, therefore, suggests that, even though previous research has found that using silent gesture limits structural transfer from participants' native language (Goldin-Meadow et al., 2008), this methodology does not completely eliminate this type of transfer (Jaffan et al., 2020). Additionally, we found no clear evidence for a preference for selecting natural orders: the natural condition was not reproduced more accurately than other conditions, and natural orders for individual dependents were not used more frequently in the majority order conditions.

What we did see evidence for was that participants generally reduced the unpredictable variation found in the training data. Moreover, they did this is a number of ways. First, our analysis of overall entropy showed that participants tended to eliminate/reduce the use of one of the competing gesture orders, so that they used one order more often across the full system. As noted above, this is similar in some ways to harmony: if participants reduce variation in word order across the system, they might produce a language that is closer to harmonic than the input language. For example, if participants in the non-harmonic natural conditions maintained the input order for adjectives, and shifted the order used for genitives away from the input pattern, then the system is not harmonic, but it is moving towards harmony. Additionally, the Mutual Information analysis showed that participants also reduced unpredictability by conditioning the word order they used based on the meaning they were trying to convey. This finding is broadly consistent with previous work, showing that people tend to reduce variation by either eliminating competing variants (here, gesture orders) or conditioning the variants on some aspect of the system (often related to meaning; see Ferdinand et al., 2019; Smith & Wonnacott, 2010; Smith et al., 2017). Together, this pattern of results suggests that while we see no impact of category-specific biases on participants' learning behaviour, nor do we see a strong preference for harmony in the strict sense in which we have defined it here, we do see evidence that participants make their output systems more regular. This is evidence that the task tapped into a system-wide bias for reducing unpredictable variation. Like harmony, other mechanisms for reducing variation in principle may involve generalisations of a pattern across aspects of a system (Culbertson & Kirby, 2016). Although it's worth noting that there may also be additional mechanisms that drive regularity that do not (see work on priming; Fehér et al., 2016).

While the evidence for the reduction of variation is in line with a number of previous studies, there is less prior work investigating whether naturalness influences learning. However, two previous studies have found some evidence for this. Do et al. (2022) examined regularisation behaviour for prenominal and postnominal adjectives in a gesture learning task. They found that participants were more likely to extend a majority gesture order signifying adjective meanings if that order was postnominal, rather than prenominal. Despite the parallels between Do et al.'s (2022) study and ours, there are also some notable differences that might explain the contrast in findings. The meanings in their study include nominals modified by adjectives, but also verb-like meanings, such as 'wave', combined in a single utterance. In addition, that study uses videos rather than still images as the prompt for meaning. It is possible that the visual parallels between the movements in the stimuli videos and the gestural videos activated the postnominal preference more strongly in their task than the ones we employed here. Little direct research has compared the effect of stimuli medium (videos vs. images) on the behaviour of participants in silent gesture experiments, but at least one study has found that even small variations in stimuli can affect participants' gesture preferences (Kirton et al., 2021). An interesting avenue for future investigation might, therefore, involve tasks where all three elements (N, Adj and Gen) are co-present and need to be ordered as a unit, as this would increase the parallel between our design and that of Do et al. (2022), as well as mirror the use of complex noun phrases in more natural language contexts.[12]

Motamedi et al. (2022) observed that, in a similar task to the one used in this study, category-specific biases for using SOV order for extensional events and SVO order for intensional events influenced regularisation behaviour. Systems which followed the preferred order patterns in their study were more likely to be systematically regularised. The difference between that study and ours may point to weaker category-specific ordering preferences for adjectives and genitives than for event-type conditioned ordering preferences. Some tentative support for this with regards to adjectives can be seen in the learning behaviour of participants in the harmony study by Culbertson et al. (2012). There, participants who were trained on a majority non-harmonic system with postnominal adjectives showed better learning than participants trained on a non-harmonic system with prenominal adjectives (Culbertson et al., 2012). The authors proposed that this pattern might be driven by a specific bias targeting adjective ordering, as we also suggest here. This postnominal adjective preference did not replicate consistently in all the subsequent follow-ups targeting different populations (Culbertson & Newport, 2015, 2017; Culbertson et al.,

[12] We thank an anonymous reviewer for highlighting this possibility.

2020a), suggesting that the effect of this category-specific bias is difficult to capture in learning tasks. However, it is worth noting that the basic word order bias in Motamedi et al. (2022), originally reported by Schouwstra and de Swart (2014), is not regularly observed in the typology of spoken languages (although see Flaherty et al., 2018; Napoli et al., 2017, for evidence in sign languages). Whereas the ordering preferences for adjectives and genitives do seem to be observable in the typology of both spoken and signed languages (Coons, 2022; but see note 1).

The distinction between these two patterns – the one observed for basic word order and the one for noun phrase dependents – could be caused by differing communicative pressures to regularise basic word orders versus harmonising phrase structure in a language system over time. For example, during early stages of language evolution, it is possible that languages which had some combination of SVO/SOV word order were widespread (Maslova, 2000; Newmeyer, 2000), and that event-type conditioned variability between these orders within a language was common, as a result of the category-specific biases which are active during language creation (Motamedi et al., 2022). Similarly, the favoured order of postnominal adjectives and prenominal genitives may have been present for similar reasons. To explain the difference in the typology for these two patterns, one explanation might be that the communicative pressure to converge on a shared, consistent basic word order across all verb types is stronger than the harmony bias present during learning for noun phrase dependents. This could, then, lead to the present-day typological distribution, where event-type conditioned basic word order is rare (Flaherty et al., 2018; Napoli et al., 2017), but non-harmonic noun phrase order is still fairly common (Dryer, 2013a, 2013b).

In combination, the results of previous research and those presented in this study reveal a complicated view of the interaction between system-wide and category-specific biases and their effect on typological word order patterns. It seems as though category-specific biases are present most strongly in improvisation tasks, but that the bias for postnominal adjectives is sometimes present in learning contexts as well. The lack of a clear preference for natural orders in our study could be due to the fact that this task activated both system-wide and category-specific biases in a limited way, especially given the lack of a clear preference for harmonic systems. There are a number of reasons for why this could be. For example, the stimuli might not have invoked the intended categories (adjective, genitive and noun) strongly enough for the related biases to be sufficiently activated. Furthermore, the overall entropy analysis and the Mutual Information calculations revealed that participants also employed strategies other than harmonisation to reduce unpredictable variation, such as eliminating competing variants and conditioning the gesture orders based on dependent types and specific nouns. These other strategies may have been especially salient to participants, as there were only two dependents and two nouns in the stimuli sets, which made the strategy of pairing these with one of the two orders a more obvious option (e.g., compared to the study in Culbertson et al., 2020a, where adult participants learned 10 modifiers and 10 nouns).

4. Conclusion

The prevalence of typological regularities across the world's languages suggests that some language features are favoured by our cognitive system. Previous research has identified a number of possible cognitive biases that give rise to specific cross-linguistic patterns, such as the system-wide bias favouring harmonic word order during learning tasks. However, exceptions to this harmonic tendency suggest that other, more category-specific biases can compete with this general preference for harmony, causing some noun phrase dependents to deviate from the harmonic tendency. These category-specific biases have mainly been observed in improvisation tasks, yet some recent evidence suggests that they may also influence learning behaviour. We tested this possibility for two noun phrase dependents which deviate from the general harmonic trend, namely, adjectives and genitives. Using a gestural artificial language learning experiment with a regularisation design, we examined if natural systems (i.e., systems with the favoured orders for these dependents) were learned better than systems with unnatural orders. Our experiment found no better learning for natural systems, nor a clear preference for harmonic systems. Consequently, the results of this study fail to support the idea that the pattern we see in the typology is the result of recurring application of competition between category-specific and system-wide biases in successive generations of language learners. Instead, we mainly found evidence of different strategies for reducing unpredictable variation, by either extending the use of one order across the whole system or by conditioning the use of different word orders on certain categories or lexical items in the data. The locus of competition between these types of biases, thus, remains unclear, and future research is needed to establish under what conditions category-specific biases influence linguistic behaviour and if this influence extends beyond contexts of language improvisation.

5. Appendix

The table in this appendix shows the type of phrase that participants produced in the translation trials at the end of the experiment. Note that it is our view that the translation data that participants provided as part of the post-test questionnaire are not likely to be very informative as to the structures that participants represented during the main experimental task. Crucially, these translations are prompted by the gesture videos, not the images, which are the stimuli used to invoke the intended meanings throughout the experimental task. Furthermore, the order of elements in the gesture videos themselves may limit the type of structure that participants produce in the translation task. For example, participants shown a postnominal gesture order for a descriptive meaning are highly likely to produce prepositional structure like “hat with spots”, simply because they try to parallel the order in the video prompt, and so a normal English NP is not produced even if that may have been how they represented the original image meaning during the experiment.

5.1 Translation data

Table 10 shows the type of phrase that participants produced in the translation trials at the end of the experiment. The coding scheme (preregistered) for this data differed between translations of adjective and genitive meanings. For genitive meanings the categories were *verb phrase* (these include possessive verb phrases like “the cyclops has a book” and action-based phrases like “the vampire wears a hat”), *genitive* (including both the *s* possessive and *of* possessive), *prepositional phrase* (including “cyclops with a cup” and “a hat on a vampire”) and *other* (including all phrases which could not be categories as any of the previous categories, such as the use of plain juxtaposition “hat vampire”). For adjective meanings the categories were *adjective* (including phrases like “spotted hat” and “stripy book”), *prepositional phrase* (like “hat with spots” or “stripes on a book”), *verb phrase* (including “the cup has spots”), and *other* (including all phrases which could not be categories as any of the previous categories, such as the use of plain juxtaposition “scarf stripes”). While the comparatively low number of pure genitive translations that participants provided may seem to undermine the idea that participants activated a possessive representation for those stimuli items, there were many instances of possessive verb phrases within the high number of verb phrase responses to genitive translation trials. Additionally, while these translations acted as an attempt to monitor the meanings that were activated in participants’ minds while participating in this task, the act of translating these meanings into English adds an additional step of transformation to those representations. Therefore we cannot take these translations as true reflections of participants’ representations of these meanings during the experimental tasks.

Table 10: Classification of English translations given for target meanings for the experiment. Note that each participant provided four translations, two for adjective meanings and two for genitive meanings. This makes for a total of 800 translations across all participants.

Class	***N* translations**
Adjective	228
Genitive	19
Preposition	191
Verb phrase	111
Other	251
Total	800

Data accessibility statement

The data collected for this study, along with all analysis and experimental materials, is openly available on OSF and can be found here.

Ethics and consent

The experimental research presented in this article was approved by the School of PPLS Research Ethics Committee at the University of Edinburgh. Informed consent to participate in the study was obtained from all participants prior to study participation.

Acknowledgements

Author AH was supported by The School of Philosophy, Psychology and Language Sciences' PhD Scholarship at the University of Edinburgh. The research was also partly supported by ESRC grant (no: ES/R011869/1) awarded to author JC, as well as Horizon 2020 grant (no: 757643) awarded to author SK. We thank the audience at the CLE talk series and the EvoLang conference for their helpful feedback and comments, as well as Elizabeth Pankratz for her excellent stats advice.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

AH conceptualised and designed the study with supervisory support from JC and SK. AH conducted all data collection, analysis and writing of the original paper draft, with JC and SK giving feedback on each draft of the paper and at each stage of the review process.

ORCiD IDs

Annie Holtz: 0000-0002-4325-9931

Jennifer Culbertson: 0000-0002-1737-6296

Simon Kirby: 0000-0002-6496-1340

References

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. *The Evolution of Language: 3rd International Conference*, 45–48.

Coons, C. (2022). Nominal word order typology in signed languages. *Frontiers in Communication*, *6*, Article 802596. https://doi.org/10.3389/fcomm.2021.802596

Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass*, *6*(5), 310–329. https://doi.org/10.1002/lnc3.338

Culbertson, J., Franck, J., Braquet, G., Barrera Navarro, M., & Arnon, I. (2020a). A learning bias for word order harmony: Evidence from speakers of non-harmonic languages. *Cognition*, *204*(2), Article 104392. https://doi.org/10.1016/j.cognition.2020.104392

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

Culbertson, J., & Newport, E. L. (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, J., & Newport, E. L. (2017). Innovation of word order harmony across development. *Open Mind*, *1*(2), 91–100. https://doi.org/10.1162/OPMI_a_00010

Culbertson, J., Schouwstra, M., & Kirby, S. (2020b). From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3), 696–717. https://doi.org/10.1353/lan.2020.0045

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

Dahl, Ö. (2008). Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua*, *118*(2), 141–150.

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), Article 5351. https://doi.org/10.21105/joss.05351

Do, M. L., Kirby, S., & Goldin-Meadow, S. (2022). Regularization of word order in the verb phrase differs from the noun phrase: Evidence from an online silent gesture perception paradigm.

In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzon (Eds.), *Proceedings of the 44th annual conference of the Cognitive Science Society* (Vol. 44).

Dryer, M. S. (2013a). Order of adjective and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013b). Order of genitive and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

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

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158–180.

Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, *184*, 53–68. https://doi.org/10.1016/j.cognition.2018.12.002

Flaherty, M., Schouwstra, M., & Goldin-Meadow, S. (2018). Do we see word order patterns from silent gesture studies in a new natural language. In T. Verhoef, C. Cuskley, H. Little, A. Ravignani, M. Flaherty, & L. McCrohon (Eds.), *The evolution of language: Proceedings of the 12th international conference* (pp. 125–127). Nicolaus Copernicus University Press. https://doi.org/10.12775/3991-1.030

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–9168. https://doi.org/10.1073/pnas.0710060105

Hartung, K., Jäger, G., Gröttrup, S., & Georges, M. (2022). Typological word order correlations with logistic brownian motion. In E. Vylomova, E. Ponti, & R. Cotterell (Eds.), *Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP* (pp. 22–26). https://doi.org/10.18653/v1/2022.sigtyp-1.3

Hawkins, R. (1981). Towards an account of the possessive constructions: NP's N and the N of NP1. *Journal of Linguistics*, *17*(2), 247–269.

Holtz, A., Kirby, S., & Culbertson, J. (2025). With or without a system: How category-specific and system-wide cognitive biases shape word order. *Cognitive Science*, *49*, e70139. https://doi.org/10.1111/cogs.70139

Jaffan, J., Klassen, G., Yang, J., & Heller, D. (2020). Semantic influences on emergent preferences of word order: Evidence from silent gesture. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual meeting of the cognitive science society* (pp. 3026–3032). Cognitive Science Society.

Jäger, G., & Wahle, J. (2021). Phylogenetic typology. *Frontiers in Psychology*, Article 2852. https://doi.org/10.3389/fpsyg.2021.682132

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245. https://doi.org/10.1073/pnas.0608222104

Kirton, F., Kirby, S., Smith, K., Culbertson, J., & Schouwstra, M. (2021). Constituent order in silent gesture reflects the perspective of the producer. *Journal of Language Evolution*, *6*(1), 54–76. https://doi.org/10.1093/jole/lzaa010

Levshina, N., Namboodiripad, S., Allassonnière-Tang, M., Kramer, M., Talamo, L., Verkerk, A., Wilmoth, S., Rodriguez, G. G., Gupton, T. M., Kidd, E., et al. (2023). Why we need a gradient approach to word order. *Linguistics*, *61*(4), 825–883. https://doi.org/10.1515/ling-2021-0098

Maslova, E. (2000). A dynamic approach to the verification of distributional universals. *Linguistic Typology*, *4*(3), 307–333. https://doi.org/10.1515/lity.2000.4.3.307

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*(2), 188–230.

Motamedi, Y., Wolters, L., Naegeli, D., Kirby, S., & Schouwstra, M. (2022). From improvisation to learning: How naturalness and systematicity shape language evolution. *Cognition*, *228*, 105206. https://doi.org/10.1016/j.cognition.2022.105206

Motamedi, Y., Wolters, L., Naegeli, D., Schouwstra, M., & Kirby, S. (2021). Regularisation, systematicity and naturalness in a silent gesture learning task. In T. Fitch, C. Lamm, H. Leder, & K. Teßmar-Raible (Eds.), *Proceedings of the 43rd annual meeting of the cognitive science society* (pp. 1451–1457, Vol. 43). Cognitive Science Society.

Motamedi-Mousavi, Y. (2017). *Artificial sign language learning: A method for evolutionary linguistics* [Doctoral dissertation, University of Edinburgh]. Edinburgh Research Archive.

Napoli, D. J., Spence, R. S., & de Quadros, R. M. (2017). Influence of predicate sense on word order in sign languages: Intensional and extensional verbs. *Language*, *93*(3), 641–670. https://doi.org/10.1353/lan.2017.0039

Newmeyer, F. J. (2000). On the reconstruction of "proto-world"word order. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 372–388). Cambridge University Press.

R Core Team. (2013). *R: A language and environment for statistical computing* [manual]. R Foundation for Statistical Computing.

Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language*, 613–644.

Rosenbach, A. (2008). Animacy and grammatical variation – findings from English genitive variation. *Lingua*, *118*(2), 151–171.

Rubio-Fernandez, P., Wienholz, A., Ballard, C. M., Kirby, S., & Lieberman, A. M. (2022). Adjective position and referential efficiency in American Sign Language: Effects of adjective semantics, sign type and age of sign exposure. *Journal of Memory and Language*, *126*, 104348.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436. https://doi.org/10.1016/j.cognition.2014.03.004

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Silverstein, M. (1986). Hierarchy of features and ergativity. *Features and Projections*, *25*, 163–232.

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1711), Article 20160051. https://doi.org/10.1098/rstb.2016.0051

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. https://doi.org/10.1016/j.cognition.2010.06.004

Strunk, J. A. J. (2004). *Possessive constructions in modern Low Saxon* [Unpublished Master's thesis, Stanford University].

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*(3), 318–330.

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535. https://doi.org/10.1073/pnas.1523631113

van Bergen, G. (2011). *Who's first and what's next. Animacy and word order variation in Dutch language production* [Doctoral dissertation, Radboud University]. Radboud Repository.

Wang, F., Kirby, S., & Culbertson, J. (2023). The learning bias for cross-category harmony is sensitive to semantic similarity: Evidence from artificial language learning experiments. *PsyArXiv*. https://doi.org/10.31234/osf.io/b3865

Weisleder, A., & Fernald, A. (2009). Real-time processing of postnominal adjectives by Latino children learning Spanish as a first language. *Proceedings of the 24th Annual Boston University Conference on Language Development*, 611–621.

Yamamoto, M. (1999). *Animacy and reference: A cognitive approach to corpus linguistics* (Vol. 46). John Benjamins Publishing.