Title
The Influence of Category-specific and System-wide Preferences on Cross-Linguistic Word Order Patterns

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Holtz, Annie
Kirby, Simon
Culbertson, Jennifer

Publication Date
2022

Peer reviewed

The Influence of Category-specific and System-wide Preferences on Cross-Linguistic Word Order Patterns

Annie Holtz (annie.holtz@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh, 3 Charles Street
Edinburgh EH8 9AD, UK

Simon Kirby (simon.kirby@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh, 3 Charles Street
Edinburgh EH8 9AD, UK

Jennifer Culbertson (jennifer.culbertson@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh, 3 Charles Street
Edinburgh EH8 9AD, UK

Abstract

Typological data shows a tendency for languages to exhibit harmonic (i.e. consistent) ordering between heads and dependents. However, some categories seem to contradict this tendency. Here we investigate one such case, the order of the noun with respect to two dependents—adjectives, which tend to follow the noun and genitives which precede. We report two silent gesture experiments examining (i) whether there are cognitive biases favouring postnominal adjective and prenominal genitive order in a single trial judgement task, and (ii) if those preferences continue to influence order when participants learn a complete word order system. Our results shed light on how biases for individual categories of elements interact with biases that affect the wider linguistic system. While participants strongly prefer postnominal adjectives and prenominal genitives when these are judged in isolation, when they learn a system of ordering, these biases are obscured and (at least in some cases) harmony emerges.

Keywords: cognitive bias; silent gesture; harmony; word order; learning

Introduction

The extent to which typological regularities are a reflection of cognitive constraints operating at the level of the individual, or cognition-external factors, remains an open question in linguistics. While some argue that typological regularities are mainly the result of lineage-specific trends (e.g., Dunn, Greenhill, Levinson, & Gray, 2011) or processes of language change (e.g., Bybee, 2008; Collins, 2019), others argue that at least some regularities are caused by cognitive biases in favour of, or against, certain language structures (e.g., Culbertson, Smolensky, & Legendre, 2012; Martin, Holtz, Abels, Adger, & Culbertson, 2020; Wilson, 2006). These biases might be active at different points in the history of a language, and during different language tasks including learning and usage (Goldin-Meadow, So, Özyürek, & Mylander, 2008; Hudson Kam & Newport, 2009; Kirby, Cornish, & Smith, 2008; Saldana, Oseki, & Culbertson, 2021).

One well-studied typological regularity in spoken languages is the cross-linguistic tendency for languages to exhibit harmonic, i.e. consistent, order between heads and dependents (Dryer, 1992; Greenberg, 1963; Hawkins, 1990). For example, languages with VO order tend to have prepositions, while languages with OV order tend to have postpositions (Dryer, 2013b, 2013e). Similarly, dependents of the same head tend to be harmonic. In noun phrases, dependents including numerals, adjectives, and demonstratives tend to be on the same side of the noun (Dryer, 2013a, 2013d).

Harmony is not a property of specific lexical items, or of specific grammatical categories, rather it is a relationship between items/categories within a system. Experimental work using artificial language learning suggest that biases like harmony are found when participants learn a system (Christiansen, 2000; Culbertson, Franck, Braquet, Barrera Navarro, & Arnon, 2020; Culbertson et al., 2012). These results show that this preference extends across phrase types and across different dependents. Harmony may be related to a bias against variability, also found in experimental work: learners tend to regularise unpredictable variation by either conditioning it on some aspect of the system or eliminating competing variants (Smith et al., 2017; Smith & Wonnacott, 2010). Harmony and regularisation can both be thought of as system-wide biases; both result in systems with less variation among different elements/categories.

While harmony appears to be a general trend, there are instances where harmony is systematically *not* present. For example, genitives (*'the child's toy'*) and adjectives (*'green grass'*) are both noun phrase dependents and, as such, one might expect harmonic ordering to prevail. However, according to WALS (Dryer, 2013a, 2013c) the order where adjectives are postnominal and genitives are prenominal (N-Adj/Gen-N = 342) is as common as the postnominal harmonic order (N-Adj/N-Gen = 342) and the prenominal harmonic order is less common (Adj-N/Gen-N = 232). Here we explore why this exception to harmony might hold. Our hypothesis is that it is due to competition between system-wide biases, like harmony, and category-specific biases that apply to individual phrases.

Previous experimental work has already uncovered some evidence for category-specific biases affecting word order. In silent gesture studies, where hearing participants must improvise gestures to convey different meanings, Culbertson, Schouwstra, and Kirby (2020) observed that participants ordered gestures denoting adjectives after the noun when pro-

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

ducing complex noun phrases. Similarly when participants produce gestures to signify the meaning of an event, they tend to produce either SOV or SVO order (Gibson et al., 2013; Goldin-Meadow et al., 2008; Hall, Mayberry, & Ferreira, 2013). Schouwstra and de Swart (2014) show that the type of event the verb denotes biases participants towards one order or the other. Verbs denoting extensional events (like *'nun throws guitar'*) tend to be expressed with SOV order, but verbs denoting intensional events (*'nun hears guitar'*) are expressed with SVO order. This kind of bias targets specific categories of events or dependents, and in some cases it has been found to weaken or disappear when embedded in a task where there is evidence of a linguistic system in place (Motamedi, Wolters, Schouwstra, & Kirby, 2021; Marno et al., 2015).

While these two bias types, category-specific and system-wide, have mainly been observed in separate linguistic tasks–improvisation and learning respectively–the typology and recent experimental work (Motamedi, Wolters, Naegeli, Schouwstra, & Kirby, 2021), suggest that category-specific biases may continue to influence behaviour during learning, meaning that orders which are produced during improvisation are also learned better. For adjectives and genitives the extended influence of the proposed category-specific biases which favour postnominal adjectives and prenominal genitives would thus continuously compete with the system-wide preference for harmony. This competition could then explain why the disharmonic N-Adj/Gen-N order is as common as the harmonic postnominal order for these two dependents.

The experiments in this study investigate whether this kind of category-specific bias exists for ordering of nouns with adjectives and genitives respectively. We do this using a silent gesture judgement task, where participants judge gesture orders for a single item in isolation–i.e., without any evidence about the linguistic system it is in. We then explore how such biases might interact with a bias for harmony active during the learning of a word order system–i.e., when participants must learn how different types of dependents are ordered.[1]

Experiment 1

The initial experiment tests participants' ordering preferences for genitives and adjectives in the absence of evidence of a conventionalised linguistic system.[2] We predict that participants' judgements will mirror the pattern seen in the typology, such that postnominal order is preferred for adjectives and prenominal order is preferred for genitives. We also predict that the postnominal adjective preference may be stronger than the prenominal genitive preference. The experiment was a between-subjects design with two conditions. In the genitive condition participants chose a gesture order to express ownership of an item. In the adjective condition they chose a gesture order to describe an item.

[1] All experimental design and analysis for both experiments were pre-registered on the Open Science Framework prior to data collection. experiment 1 here and experiment 2 here.

[2] The experimental setup was based on Experiment 1b in Motamedi, Wolters, Naegeli, Kirby, and Schouwstra (2021).

Methods

Materials The experiment was developed to run in participants' web browsers using the JavaScript library jsPsych (de Leeuw, 2015). Participants saw a collection of grayscale digital drawings (see figure 1 for examples) showing either instances of item ownership (genitive condition, e.g. '*vampire's hat*') or items with different patterns (adjective condition, e.g. '*striped cup*'). For each image there were two gesture videos. The videos showed a model gesturer producing two gestures in sequence, one for the head noun and one for the adjective/genitive dependent. The videos differed only in the order of these two gestures – in one the head noun was the first gesture, in the other it was the last. Each phrase component was denoted using a gesture made with both hands and the videos ended with both hands in a neutral position. The videos were all 4389 milliseconds long.

Figure 1: Sample stimuli images for genitive condition (left) and adjective condition (right).

Procedure Participants were randomly assigned to a condition. At the start of the experiment, participants were shown a sample 2x2 image grid with the kinds of images they would encounter during their test trial. After this familiarisation trial, participants were instructed that they would see the same kind of 2x2 grid again but one of the images would be highlighted. They were told that two videos would appear next to the images and that these represented two ways to express "ownership of the item" (genitive condition) or two ways that the "item could be described" (adjective condition) in a made-up sign language. Their task was to choose the gesture video which they thought best conveyed the meaning of the highlighted image. The images were displayed and the videos looped until participants chose one gesture order by clicking on one video. The next trial asked participants to indicate on a slider how strong their preference was for the gesture order they chose in the forced-choice trial. The two gesture videos looped on either side of a slider with the slider point starting in the middle. The slider was marked with "weakly prefer video a/b" and "strongly prefer video a/b" on either side of the mid-point. Following this, participants were shown the video they had chosen in the forced-choice trial and were asked to translate the meaning of the gesture video into English. Finally, participants responded to two short demographics questions. One asking them if they knew a sign language and another asking them to note which spoken languages they knew and at what level of proficiency.

Participants 384 participants were recruited through the online crowdsourcing platform Prolific. Participants were prescreened such that only people who reported English as their first language, had at least a 95% previous task approval rate, and had not completed any of our previous experiments could participate. Participants were paid the equivalent of £8.91 per hour. After excluding participants who indicated proficiency in a sign language (N=8), and participants who responded too quickly to the forced-choice trial (< 9678 milliseconds, combined time for both videos) and/or who did not indicate a preference for the same gesture order across both the forced-choice and slider trial (N=56), there were 160 participants in each condition.

Results

The proportion with which participants chose the predicted order in each condition–prenominal in the genitive condition and postnominal in the adjective condition–closely parallels what is seen in the typological data (see figure 2). Responses were analysed using mixed effects logistic regression models implemented using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2013) unless otherwise noted. Two intercept-only models indicated that participants chose the predicted order for their respective conditions at rates significantly above chance (genitive: $\beta = 0.56$, SE = 0.16, z = 3.43, $p < 0.001$, adjective: $\beta = 0.51$, SE. = 0.16, z = 3.02, $p < 0.01$).[3]

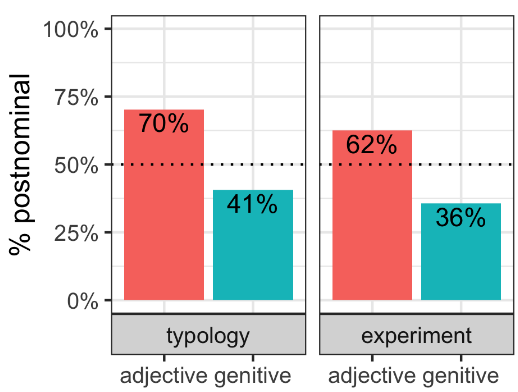

Figure 2: Proportion of postnominal orders per dependent type in both the typological and experimental data. In both the typology and the experiment, postnominal order is preferred for adjectives, prenominal order for genitives.

In the typology, the postnominal adjective preference appears to be stronger than the prenominal genitive preference. There was no evidence for this in the experiment. A logistic regression model including condition as a fixed effect revealed no difference between the two conditions ($\beta = 0.05$, SE. = 0.23, z = 0.23, p = 0.82).[4]

We ran two linear models on the slider data from each condition. These models tested whether participants who chose the predicted order in the forced-choice task tended to give these videos a higher rating in the slider task, compared to those who did not choose the predicted order.[5] Neither model reached significance, suggesting that preference ratings were equally strong across both of these groups (genitive: $\beta = 3.89$, SE = 2.14, t = 1.81, p = 0.071, adjective: $\beta = 1.06$, SE. = 2.07, t = 0.51, p = 0.61). Finally we ran a linear model which included an interaction between the fixed effects of predicted order and condition. The results revealed no significant interaction between these effects, confirming that there was no difference in preference ratings for participants who chose the predicted order across both conditions. However, there was a significant positive coefficient for condition ($\beta = 4.93$, SE = 2.36, t = 2.09, $p < 0.05$) indicating that slider ratings were, overall, slightly higher in the adjective condition.

Discussion

The results from experiment 1 shown that the orders participants prefer to use for descriptive and possessive meanings are the most common orders we see for adjectives and genitives in typology. These results suggest that the unexpected absence of a preference for harmonic alignment of these two orders in the typology may in part be a reflection of these category-specific preferences. In the next experiment, we explore whether these biases persist when participants are taught a miniature language system in which the order of both dependent types must be learned. As mentioned above, it is under these circumstances that we expect harmony to have an influence (e.g., as in Culbertson et al., 2012) and thus to interact (or compete) with category-specific biases.

Experiment 2

Experiment 2 explored how the category-specific biases identified in experiment 1 interact with a bias for harmony in a task where participants must learn how different types of dependents are ordered. Experiment 2 was a between-subjects design based on Motamedi, Wolters, Naegeli, Schouwstra, and Kirby (2021), Ferdinand, Kirby, and Smith (2019), and Culbertson et al. (2012) which use a regularisation design. In regularisation experiments participants are trained on variable input and, at test, the extent to which they regularise or retain the input variability is measured. The idea being that participants will regularise preferred variants more reliably than

[3]The model for the adjective condition includes a random effect for items (i.e., the picture conveyed). The model did not converge in the genitive condition, therefore a logistic regression model with no random effect (for item) is reported here.

[4]This could reflect some influence from native language: English tends to have prenominal genitives for the kinds of possessive meanings used in the experiment (*vampire's hat)* whereas adjectives are rarely postnominal. Thus the postnominal adjective preference competes with the native language influence in the adjective condition, whereas in the genitive condition participants' native language is (at least partly) consistent with the proposed underlying preference. On the other hand, it is difficult to determine whether the typological difference, here based on raw counts of languages in the sample, is reliable.

[5]Neither model converged with random intercepts for item.

dis-preferred variants.

The experiment had four conditions and the manipulation between conditions was the proportion of prenominal vs postnominal gesture orders that participants observed in training. Each condition had a majority gesture order and a minority gesture order. In the natural condition the majority orders for both dependents were those preferred by participants in experiment 1 (75% N-Adj & Gen-N) and in the unnatural condition the majority orders were the reverse (75% Adj-N & N-Gen). In the majority prenominal condition the majority order for both dependent types was prenominal (75% Adj-N & Gen-Adj) and in the majority postnominal condition the majority order was always postnominal (75% N-Adj & N-Gen).

Methods

Materials This experiment was developed using jsPsych and used the same stimuli as in experiment 1.

Procedure Participants were randomly assigned to one of the four conditions outlined above and a pseudo-randomised stimuli set. The stimuli set consisted of two nouns, each of which were randomly assigned to appear with one of the genitives and one of the adjectives. At the start, similar to experiment 1, participants were familiarised with the stimuli via a grid displaying the adjective and genitive images in their stimulus set. During the training phase, participants saw these same image grids with one image highlighted. This was accompanied by a gesture video playing under the images. Each of the four target images was highlighted eight times (32 trials total). In six of these eight exposures the gesture video showed the majority order for that condition and twice it showed the minority order. Participants were told to sit back and watch the images and videos. The trials progressed automatically. In the testing phase, participants again saw an image grid with one image highlighted accompanied by both possible gesture videos. Participants were instructed to "click on the corresponding gesture video" like they saw during training. Participants saw each target image eight times (32 trials total) and clicked a centred "Next" button between each trial. After the testing phase participants provided a translation for all four target meanings (prompted by a gesture video) and answered the same demographic questions as in experiment 1.

Participants 211 participants were recruited using the same platform and criteria from experiment 1. Participants were paid the equivalent of £8.91 per hour. Participants who had knowledge of a sign language (N=8) and who continuously clicked the same button in testing (N=3) were excluded. The final number of participants was 50 in the natural, 53 in the unnatural, 46 in the majority prenominal, and 51 in the majority postnominal condition.

Results

Our main predictions for this experiment were that participants would be able to learn the orders they were trained on (i.e. reproduce or over-produce the majority variant they were trained on) and that there would be an effect of naturalness, such that participants would select more natural (preferred) orders than what is predicted by chance, and that they would learn the natural orders more readily than the unnatural orders. Furthermore, we predicted that participants would reduce the variability in their input in two ways, by selecting one order for a given dependent type more consistently (regularisation, measured as change in conditional entropy between input and output) and by selecting the same order across both dependent types (harmony, measured as change in entropy between input and output).

The overall distribution of orders that participants selected across all conditions can be found in figure 3.

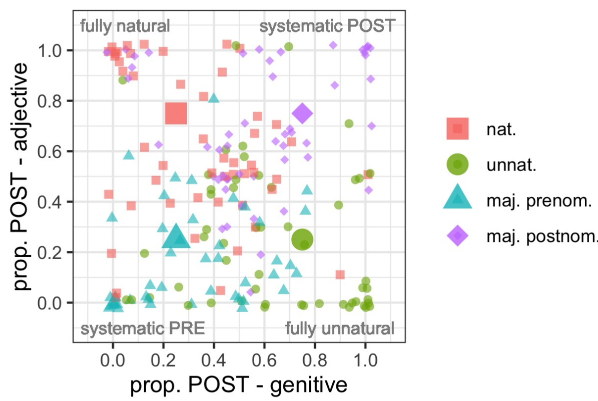

Figure 3: Proportion of postnominal orders selected by each participant for both dependent types. The four larger shapes represent the input proportions per condition.

Learning The overall proportion of majority orders selected by participants for each condition is shown in figure 4 (top). To test whether use of the majority input order differed across condition, we ran a mixed effects logistic regression model with condition and dependent type, as well as their interaction as fixed effects. We included a by-participant random intercept, with a slope for dependent type. The outcome variable was binary: 1 indicated a match between training majority order and video chosen and 0 indicated a mismatch. Both condition and dependent type were deviation coded. The model had a significant positive intercept (β = 1.34, SE = 0.12, z = 11.24, $p < 0.001$), showing that, on average, participants across all conditions choose the majority order at a rate above chance. There was also an effect of condition, such that participants in the majority prenominal condition chose the majority order more often, compared to the grand mean (β = 0.64, SE = 0.21, z = 3.05, $p < 0.01$). Furthermore, participants also selected the majority order *less* often for the genitive meanings (β = -0.74, SE = 0.18, z = -4.24, $p < 0.001$), except in the natural condition, where they selected the majority order for genitives *more* often (β = 0.99, SE = 0.29, z = 3.40, $p < 0.001$), compared to the grand mean.

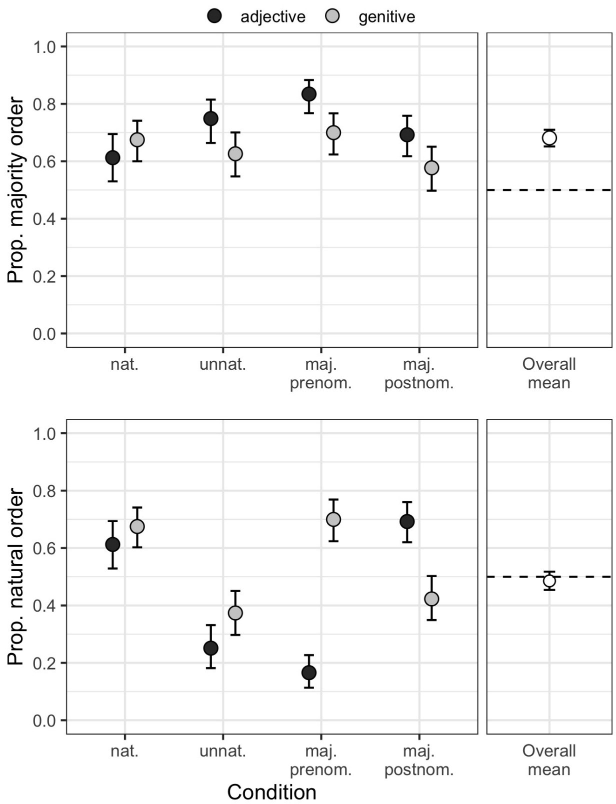

Figure 4: Top: Proportion of test trials where participants selected the majority input order for each condition and dependent type. Right-hand facet shows grand mean. Participants tended to reproduce the majority orders from their training. Bottom: Proportion of test trials where participants selected the natural order (i.e., prenominal for genitives, postnominal for adjectives) for each condition and dependent type. Right-hand facet shows the grand mean. There was no overall preference for natural orders. All error bars represent bootstrapped 95% confidence intervals.

Naturalness The overall proportion of natural orders selected across conditions is shown in figure 4 (bottom). Recall that the predicted natural orders are postnominal for adjectives and prenominal for genitives. To test whether use of the natural order differed across conditions, we ran a logistic mixed effects model using the binary dependent variable Natural (1 signifying a match between the predicted natural order and the selected order, 0 meaning a mismatch). The rest of the model structure was identical to the one used for learning above. The model did not show a significant intercept (as suggested by figure 4). However, selection of the natural order was more likely in the natural and majority postnominal conditions (natural: $\beta = 1.14$, SE = 0.18, z = 6.49, $p < 0.001$, postnominal: $\beta = 0.59$, SE = 0.17, z = 3.36, $p < 0.001$), compared to the grand mean. Conversely, participants in the unnatural and prenominal conditions selected the natural order *less* often compared to the grand mean (unnatural: $\beta = -1.24$, SE = 0.17, z = -7.11, $p < 0.001$, prenominal: $\beta = -0.88$, SE = 0.18, z = -2.62, $p < 0.01$). Finally, there were significant interactions between dependent type and condition. For the majority prenominal condition, participants used the natural order more for genitives, likely since it is consistent with the majority order ($\beta = 3.03$, SE = 0.36, z = 8.40, $p < 0.001$). For the majority postnominal condition participants used the natural order less often for genitives, since this conflicted with the majority order ($\beta = -2.71$, SE = 0.34, z = -7.94, $p < 0.001$).

Harmony A fully harmonic language refers to a system in which all meanings are expressed with the same gesture order. As such, harmony constitutes a reduction in overall system variation. We quantify an increase in harmony as an overall reduction in entropy, following Ferdinand et al. (2019) and Motamedi, Wolters, Naegeli, Schouwstra, and Kirby (2021). The entropy (H) of a system is defined as:

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

where, (V) refers to the two gesture variants (prenominal and postnominal orders). 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). The remaining two conditions have a training entropy of 0.8112781. We measured the change in entropy between these input values and the orders participants selected to see if participants tended to reduce variation by harmonising the system. Figure 5 (top) shows the mean change in entropy for each condition as well as the overall mean change.

To evaluate if these changes are reliably greater than zero we calculated bootstrapped confidence intervals around the reported mean entropy changes for each condition.[6] These were generated using the 'boot' package in R and based on 10,000 samples. Our findings suggest that all but one of the conditions—majority postnominal—show a reliable reduction in entropy.[7] Furthermore, confidence intervals around differences between conditions reveals 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).

Regularisation Another way in which variability in a system can become more consistent is by reducing the variants (orders) used for a particular type of dependent. The result is not harmony, but a more regular pattern for a given dependent type. We refer to that as regularisation, and measure it using

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

[7] The lack of harmonising behaviour for participants in this condition could be due to participants' native language. The majority orders in this condition are the opposite to those used most commonly in English for these dependents.

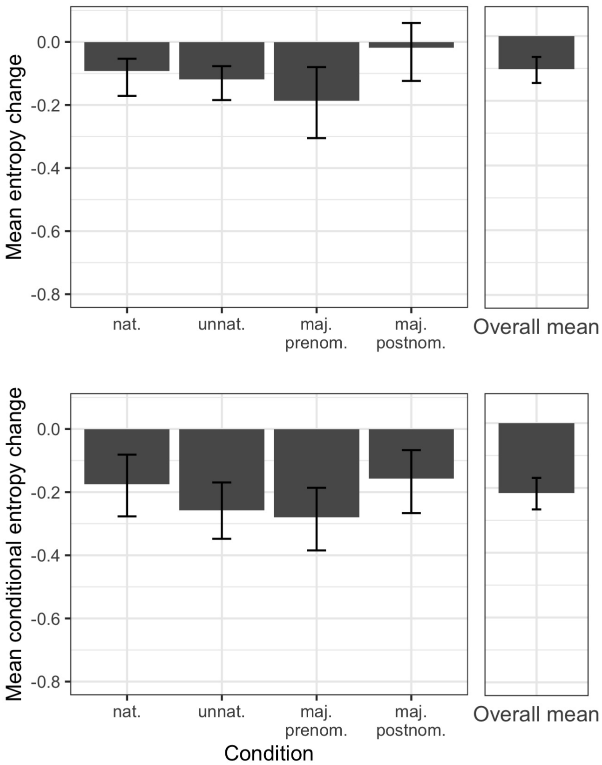

Figure 5: Top: Mean change in entropy between input and output, showing participants tend to harmonise, except in the postnominal condition. Right-hand facet shows grand mean change in entropy. Bottom: Mean change in conditional entropy, showing tendency to regularise in all conditions. Right-hand facet shows grand mean change in conditional entropy. Error bars represent bootstrapped 95% confidence intervals.

conditional entropy:

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

Change in conditional entropy is calculated per participant by comparing the input entropy (0.8112781 across conditions) and the output entropy (see figure 5, bottom). The 95% confidence intervals for the mean of each condition (figure 5, bottom) and the difference between means across conditions were calculated in the same way as for our harmony measure. The results suggest that conditional entropy decreases across all conditions, and that there is no reliable difference between conditions in terms of this change.

Discussion

The experiments reported here examined two things, the presence of category-specific word order preferences for adjectives and genitives and how these preferences compete with a system-wide bias towards harmonic word order.

Experiment 1 showed that, in the absence of a linguistic system, participants preferred gesture orders that were consistent with typological word order tendencies: postnominal order for adjectives and prenominal for genitives. Experiment 2 found that when participants had to learn an ordering system, evidence for the preferences identified in experiment 1 was largely absent. Instead, there was evidence for system-wide biases: participants tended to regularise the order they used for a given dependent type, or harmonise the language such that they used a single order more consistently across dependent types. The latter results are broadly consistent with work showing that people tend to reduce input variation, either by conditioning the variation on some aspect of the system or by reducing the number of variants used (Smith et al., 2017; Smith & Wonnacott, 2010).

The fact that the category-specific word order preferences do not seem to influence participants' learning behaviour in experiment 2 contrasts with at least one previous study showing that this kind of bias may be active during learning. Motamedi, Wolters, Naegeli, Schouwstra, and Kirby (2021) found that a preference for using SOV for extensional events and SVO for intensional events identified in a single-trial judgement task (i.e., where no linguistic system is evident) also influenced behaviour in a learning task very similar to the one used here. It is possible that the preferences in this task are relatively weaker and thus do not survive in competition with system-wide biases in a learning task.[8] However, the bias found in Motamedi, Wolters, Naegeli, Schouwstra, and Kirby (2021) (originally reported by Schouwstra & de Swart, 2014) is not clearly found in the typology of spoken languages (although see Flaherty, Schouwstra, & Goldin-Meadow, 2018; Napoli, Spence, & de Quadros, 2017, for evidence in sign languages). By contrast, we do see traces of the two category-specific biases found in experiment 1 across both language types (Coons, 2022).

Another possible explanation for the lack of category-specific preferences in experiment 2 is failure of the task to activate the intended dependent categories. In order for participants to display category-specific preferences the relevant categories need to be accessed by participants. It is possible that, when participants were faced with a more complex system in experiment 2, these categories were not activated as clearly as in experiment 1. Future work will examine if a task involving both improvisation and learning might prove more successful at activating these categories, while still allowing us to examine competition between these category-specific biases and the system-wide bias for harmony.

[8] A conceptual parallel may exist between the event-type results and those found here. A preference for SVO for intensional events could result from the fact that in these events the existence of the objects depends on the action of the verb (e.g. *'gnome dreams of banana'*), whereas in extensional events the object exists independently from the verb (e.g. *'nun throws guitar'*). Adjectives, like intensional events, may denote properties which depend on the item for their interpretation (e.g. *'tall building'*). Conversely, possessors (genitives) are more independent of the objects they possess. Thus the deeper semantic motivation for postnominal adjectives and prenominal genitives may be related to the motivation for the intensional-extensional ordering preferences.

Acknowledgments

This project was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757643). We wish to thank the anonymous reviewers for their thoughtful comments, as well as Yasamin Motamedi, Alexander Martin, Mora Maldonado, and the audience at the CLE talk series for helpful feedback and discussions.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bybee, J. L. (2008). Formal universals as emergent phenomena: The origins of structure preservation. In J. Good (Ed.), *Linguistic universals and language change.* New York: Oxford University Press.

Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In *The evolution of language: Proceedings of the 3rd international conference* (pp. 45–48). Paris: Ecole Nationale Superieure des Telecommunications.

Collins, J. (2019). Some language universals are historical accidents. In K. Schmidtke-Bode, N. Levshina, S. M. Michaelis, & I. Seržant (Eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence.* Berlin: Language Science Press.

Coons, C. (2022). Nominal word order typology in signed languages. *Frontiers in Communication*, *6*, Article 802596.

Culbertson, J., Franck, J., Braquet, G., Barrera Navarro, M., & Arnon, I. (2020). A learning bias for word order harmony: evidence from speakers of non-harmonic languages. *Cognition*, *204*, Article 104392.

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3), 696–717.

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

de Leeuw, J. (2015). jsPsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, *68*(1), 81–138.

Dryer, M. S. (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.

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

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

Dryer, M. S. (2013d). 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.

Dryer, M. S. (2013e). Order of object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: 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.

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

Flaherty, M., Schouwstra, M., & Goldin-Meadow, S. (2018). Do we see word order patterns from silent gesture studies in a new natural language. In *The evolution of language: Proceedings of the 12th international conference.* (pp. 125–127). Torun: Nicolaus Copernicus University Press.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*, *24*(7), 1079–1088.

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.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language.* Cambridge, MA: MIT Press.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry*, *21*(2), 223–261.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Marno, H., Langus, A., Omidbeigi, M., Asaadi, S., Seyed-Allaei, S., & Nespor, M. (2015). A new perspective on word order preferences: the availability of a lexicon triggers the use of svo word order. *Frontiers in Psychology*, *6*, Article 1183.

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa: a journal of general linguistics*, *5*(1), 97-116.

Motamedi, Y., Wolters, L., Naegeli, D., Kirby, S., & Schouwstra, M. (2021). From improvisation to learning: how naturalness and systematicity shape language evolution. *PsyArXiv*.

Motamedi, Y., Wolters, L., Naegeli, D., Schouwstra, M., & Kirby, S. (2021). Regularisation, systematicity and naturalness in a silent gesture learning task. In *Proceedings of the 43rd annual meeting of the cognitive science society* (pp. 1451–1457). Vienna: Cognitive Science Society.

Motamedi, Y., Wolters, L., Schouwstra, M., & Kirby, S. (2021). The effects of iconicity and conventionalisation on word order preferences. *PsyArXiv*.

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.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna: R Foundation for Statistical Computing. Retrieved from `http://www.R-project.org/`

Saldana, C., Oseki, Y., & Culbertson, J. (2021). Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language*, *118*, Article 104204.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

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*, *372*(1711), Article 20160051.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982.