Human information processing shapes language change

Short title: Information processing shapes language

Maryia Fedzechkina*[,1]
Becky Chu[2]
T. Florian Jaeger[2,3,4]

[1] Department of Psychology, University of Pennsylvania,
Philadelphia, PA 19104 USA

[2] Department of Brain & Cognitive Sciences, University of Rochester,
Rochester, NY 14627-0268 USA

[3] Department of Computer Science, University of Rochester,
Rochester, NY 14627-0268 USA

[4] Department of Linguistics, University of Rochester,
Rochester, NY 14627-0268 USA

*Please address correspondence to:
Maryia Fedzechkina
Institute for Research in Cognitive Science,
3401 Walnut str, Suite 400A
Philadelphia, PA 19104 USA
Email: mfedze@sas.upenn.edu
Tel: +1 (585) 733-6080

**Abstract**

Human languages exhibit both striking diversity and abstract commonalities. Whether these commonalities are shaped by potentially universal principles of the human information processing has been of central interest in the language and psychological sciences. Research has identified one such abstract property in the domain of word order: although sentence word order preferences vary across languages, the superficially different orders result in short grammatical dependencies between words. As short dependencies are easier to process, these findings raise a possibility that languages are shaped by biases of human information processing. The current study directly tests the hypothesized causal link. We find that learners exposed to novel miniature artificial languages that have unnecessarily long dependencies systematically restructure the input to reduce dependency lengths rather than follow the surface preference of their native language, thus providing direct evidence for a causal link between processing preferences in individual speakers and patterns in linguistic diversity.

**Keywords**

Language universals; language processing; learning biases; language structure; language evolution

## 1. Introduction

Natural languages vary along many dimensions but this variation is not random—unrelated languages appear to share a striking number of underlying similarities. Understanding the constraints underlying these similarities has been the central question in the biological and language sciences as most theories agree this can shed light on the mechanisms of language processing and representation in the human brain (e.g., Bates & MacWhinney, 1982; Chomsky, 1965; Christiansen & Chater, 2008; Fodor, 2001; Givón, 1991; Greenberg, 1963; Hawkins, 2014; Slobin, 1973). Theories differ, however, in how they view the nature of the hypothesized constraints. On one view, they are specific to language acquisition and not shared by other cognitive systems (Chomsky, 1965; Fodor, 2001). Another proposal holds that cross-linguistic distributions reflect more general cognitive biases that are rooted in human information processing (Christiansen & Chater, 2008; Hawkins, 2014; Kirby, Tamariz, Cornish, & Smith, 2015)

Here we experimentally test the latter proposal on one of the most basic and perhaps most well-studied grammatical properties of human languages -- the way in which they order information in a sentence. While the order of words in a sentence varies across languages, this variability is constrained (Dryer & Haspelmath, 2011; Greenberg, 1963). We ask whether one of the factors that constrains the cross-linguistic distribution of word orders is a well-documented bias of language processing towards short *grammatical dependencies* (Demberg & Keller, 2008; Grodner & Gibson, 2005).

Grammatical dependencies are asymmetric relations between the head (a word that licenses the presence of other words) and a dependent (a word that modifies the head). For example, in the sentence 'The boy is kicking the ball', the head (verb 'kick') forms two grammatical dependencies – one with the subject ('the boy') and one with the direct object ('the ball'). Psycholinguistic evidence shows that dependency length (i.e., the distance between the head and its dependent) affects processing efficiency: longer dependencies are associated with greater processing difficulty in comprehension than shorter dependencies (Demberg & Keller, 2008; Grodner & Gibson, 2005), presumably due to increased memory demands (Bartek, Smith, Lewis, & Vasishth, 2011). Paralleling this processing advantage, language *production* also exhibits a preference for shorter dependencies. For example, when several word order choices are available to convey the same message, speakers of verb-initial languages like English tend to order constituents short-before-long (Arnold, Wasow, Losongco, & Ginstrom, 2000; Wasow, 2002), while speakers of verb-final languages like Japanese typically have a long-before-short preference (Ros, Santesteban, Fukumura, & Laka, 2015; Yamashita & Chang, 2001). The respective verb-dependent orderings reduce the average dependency length in a sentence (see Fig. 1).

Recent large-scale computational studies have shown that this preference for shorter dependencies in production is not limited to situations of meaning-equivalence. Rather, the word order distributions in human languages overall exhibit properties that reduce the average dependency length of their sentences: all languages studied so far (almost 40) have average dependency lengths that are significantly shorter than would be expected by

chance (Ferrer i Cancho, 2004; Futrell, Mahowald, & Gibson, 2015; Gildea & Temperley, 2010; Liu, 2008), with some languages being close to the theoretical minimum of dependency length (Gildea & Jaeger, under review; Gildea & Temperley, 2010).

In short, work in psycholinguistics has shown a processing advantage of shorter dependencies in comprehension and production and work in linguistic typology has documented a parallel prevalence of shorter dependencies in grammars of human languages. What *causes* this correlation? In two experiments, we test whether the cross-linguistic bias towards shorter dependencies originates in the limitations of the human processing system. Specifically, we ask whether a preference for short dependencies influences language acquisition before a mature linguistic system is in place, causing learners to produce languages that deviate slightly from the original input. This would introduce small deviations into the newly acquired language and slightly shift the input for the next generation of learners towards a grammar with shorter dependencies. If these deviations spread through the population, small deviations can accumulate over generations, causing the language to have increasingly shorter dependencies. This would explain the cross-linguistic tendency of languages to have shorter than expected average dependency lengths. It would then also follow that at least some striking cross-linguistic patterns do *not* require a reference to cognitively arbitrary linguistic-specific biases (cf. Christiansen & Chater, 2008).

We test the first step in this chain. We use a miniature artificial language learning paradigm (Hudson Kam & Newport, 2009; Kirby et al., 2015) to directly probe the causal link between processing biases in individual language learners and the preference for shorter dependencies observed cross-linguistically. Miniature language learning has been successfully used to study the mechanisms underlying first and second language acquisition (Pajak & Levy, 2014; Saffran, Aslin, & Newport, 1996). Recent work has adapted this paradigm to explore the underlying causes of cross-linguistic patterns by creating situations of atypical input (reminiscent of situations of pidgin or language change) in the laboratory and studying how learners deviate from the atypical input they receive (Culbertson & Adger, 2014; Culbertson, Smolensky, & Legendre, 2012; Fedzechkina, Jaeger, & Newport, 2012; Fedzechkina, Newport, & Jaeger, 2016; Hudson Kam & Newport, 2009; Kirby et al., 2015; Smith & Wonnacott, 2010).

Here, we present learners with input languages that have inefficient (unnecessarily long) dependencies, and test whether learners shift the language towards more efficient (shorter) dependencies.

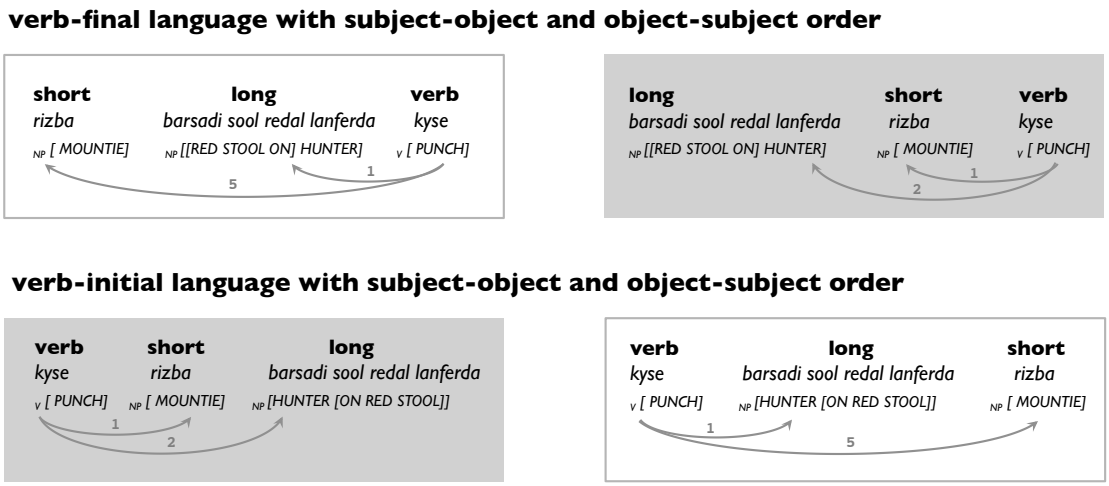


Figure 1: Comparison of dependency lengths for two possible argument orderings (subject, object) in verb-final vs. verb-initial languages. All sentences express the same meaning. Arches represent grammatical dependencies between the verb and the head of its two arguments. Numbers represent dependency lengths, measured in words. For verb-final languages (top panel), ordering dependents long-before-short leads to shorter total dependency length between the dependents and their head (the verb). For verb-initial languages (bottom panel) this relationship between the length of the dependent and its order relative to the head is reversed: here, ordering dependents short-before-long leads to shorter overall dependency length.

## 2. Method

### 2.1. Participants

Participants in the experiments were monolingual native English speakers between ages 18 and 30 recruited from the University of Rochester and the surrounding community.

Each participant took part in only one of the experiments and received $30 for its completion. Following earlier work (Fedzechkina et al., 2012; Fedzechkina et al., 2016), recruitment continued until 20 participants, who successfully learned the language, were enrolled in each experiment. Recruiting and execution of the studies was approved by the Research Subjects Review Board at the University of Rochester.

### 2.2. Design and Materials

In two experiments, monolingual native speakers of English learned a miniature artificial language by watching short videos describing simple transitive events performed by two human actors (e.g., 'chef punch referee') and hearing their descriptions in the novel miniature language. The languages in both experiments had flexible word order, so that subject-object (SO) and object-subject (OS) orders occurred equally frequently in the input. Like many languages with flexible word order (Blake, 2001), our languages had consistent case-marking – a noun suffix that disambiguated who was doing what to whom in the scene. The case-marker was always 'di' and occurred on all direct objects. The languages shared the same lexicon of 4 transitive verbs, 8 nouns (6 animate and 2 inanimate), 3 adpositions ('with', 'next-to', 'on'), and 2 color adjectives ('blue' and 'red'), see SI for more details. Both languages contained adpositional phrases (e.g., 'chef next-to blue skateboard', see Fig. 2). The order of the adposition (e.g., 'next-to') relative to its dependent ('blue skateboard') and head ('chef') followed cross-linguistically common patterns (Dryer & Haspelmath, 2011), as shown in Figure 1.

The miniature languages differed in whether they were verb-final (Experiment 1) or verb-initial (Experiment 2). The verb-final language in Experiment 1 used pre-nominal postpositional phrases (as in Japanese or Hindi), ordering the adposition after its dependent and before its head (e.g., 'blue skateboard next-to chef'). The verb-initial language in Experiment 2 used post-nominal preposition phrases (as in English), ordering the adposition after its head and before its dependent (e.g., 'chef next-to blue skateboard').

In training, participants were exposed to sentences that either contained two 'short' constituents (i.e., both subject and object without adpositional phrase modification; 50% of training scenes) or two 'long' constituents (i.e., subject and object with adpositional phrase modification; 50% of training scenes). Sentences, in which subject and object phrases differed in length were not part of the input. Word order was thus independent of phrase length in the input, and both short-short and long-long scenes occurred equally frequently with OS and SO orders. During the production test, participants described previously unseen scenes that either contained one long constituent (either subject – in 1/3 of production scenes or object – in 1/3 of production scenes) or no modification of either constituent (1/3 of production scenes).

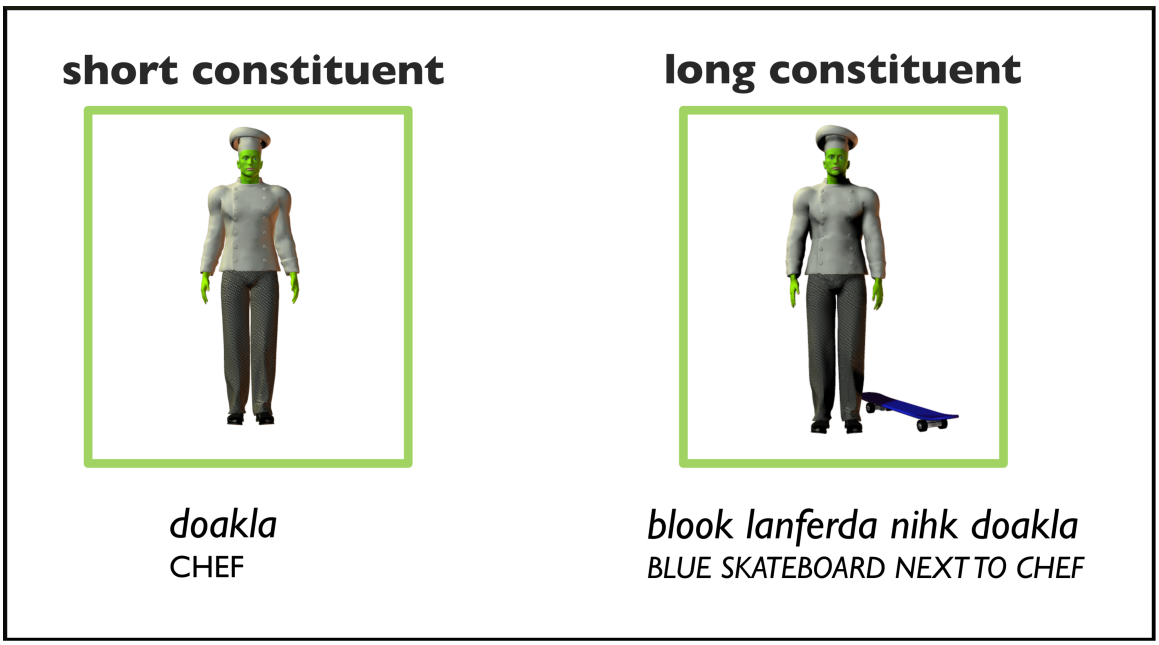


Figure 2: Illustration of constituent length manipulation in the experiments. The visual referent on the left can be described with a short phrase. The referent on the right requires a more complex long phrase. Example descriptions are shown for the verb-final miniature language from Experiment 1 (provided only auditorily in the experiment) along with their English glosses (not shown to participants).

### 2.3. Prediction

If the cross-linguistic pattern of grammars with shorter-than-expected dependencies indeed originates in processing biases of individual users, we would expect learners to introduce a preference for shorter dependencies in their productions. Specifically, dependencies are minimized in long-before-short ordering in the verb-final language (Experiment 1) and in short-before-long ordering in the verb-initial language (Experiment

2), see Fig. 1. Critically, this predicts opposite ordering preferences for the two languages, thereby making surface order-based biases originating in the learners' native language an unlikely source of our result: all participants in our experiments are monolingual native speakers of English, a verb-initial language with an overall short-before-long ordering preference (Arnold et al., 2000; Lohse, Hawkins, & Wasow, 2004; Wasow, 2002). Thus, an inverse long-before-short preference in the verb-final language, if observed, would lend support for the hypothesis that processing constraints shape language structures over time.

### 2.4. Procedure

Both experiments were conducted in 3 1-hour sessions on consecutive days with at most 1 day in between. Each session involved a similar combination of exposure and test blocks, with more intensive vocabulary exposure on day 1 and more intensive sentence exposure on days 2 and 3 (see Fig. 3).

***Noun exposure.*** Participants saw pictures of characters or objects one at a time, accompanied by their names in the novel language and were instructed to repeat the names out loud to facilitate learning.

***Vocabulary tests.*** Following noun exposure, participants completed noun comprehension and production tests. In the comprehension test, participants were shown a set of four character pictures accompanied by a name in the novel language and asked to choose the

character matching the name. In the production test, participants were asked to name the character shown on the screen. Feedback on performance was provided after each trial in both tests.

***Phrase exposure and tests.*** Participants were explicitly informed that they would learning phrases in the new language. These contained a character modified by a description (see Fig. 2 and SI for more details). The same procedure as in vocabulary training and tests described above was used here.

***Sentence exposure.*** Participants learned the grammar by watching short videos and hearing their descriptions in the novel language. Participants were instructed to repeat the sentences aloud to facilitate learning. On day 1, participants could replay the videos and the sound as many times as they wished; no repetitions were allowed on subsequent days.

***Sentence comprehension test.*** Participants were presented with two side-by-side videos accompanied by auditory descriptions. The videos showed the same action and characters but the order of the actor and patient of the action reversed. Participants were asked to choose the video that matched the description. Feedback on performance was provided on each trial.

***Production test.*** Participants were shown two side-by-side previously unseen videos depicting the same action and participants in switched roles. One of the videos was highlighted. The videos disappeared from the screen after 1200ms and were replaced by a

crosshair in the center of the screen. Participants were instructed to describe the highlighted video after seeing the crosshair. A verb prompt was provided to facilitate the descriptions. No feedback on performance was provided during this test.

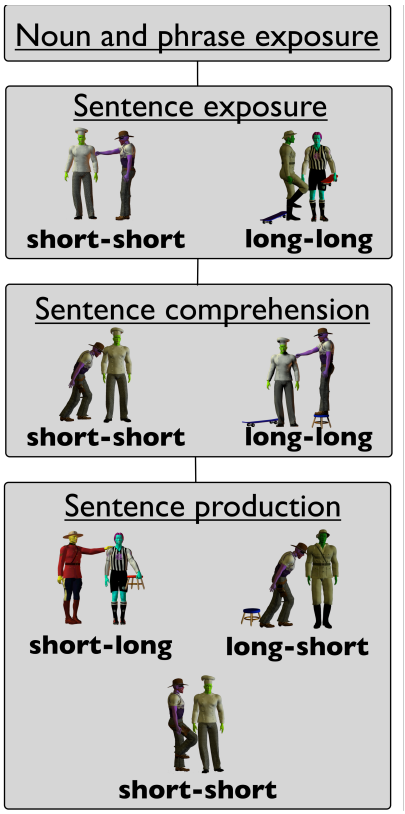


Figure 3: Each participant was exposed to a novel miniature language following the procedure shown here on each of the three separate visits to the lab (images represent still pictures of video stimuli used in the experiments). The measure of interest was the word orders learners produced in the sentence production phase.

## 4. Results

### 4.1. Accuracy of acquisition

Both experiments followed the same scoring procedure. In production, we scored word order used in the utterance overall and in the adpositional phrase (if present), the presence of case-marking on the object, lexical (using incorrect vocabulary) and grammatical mistakes (using word order not allowed by the language in a sentence overall or within an adpositional phrase, using a case-marker on a constituent other than the object). Occasionally, participants mispronounced/misused the name of a referent or an action. If it was impossible to determine the sentence word order (e.g., when both referent names were mispronounced), the production was scored as both lexically and grammatically incorrect. If it was still possible to determine the word order used in the utterance (e.g., the name of only one referent was incorrect), the production was coded as a lexical mistake and was scored for grammatical accuracy. All analyses reported here were based on grammatically correct productions only. On occasion, participants used an adpositional phrase inappropriately for the visual scene (e.g., they added it if the scene required no modification or dropped it when the scene required modification). These cases were exceedingly rare (2.2% of productions across all days of training in Experiment 1 and 3.5% in Experiment 2) and were excluded from all further analyses.

In comprehension, we scored whether participants chose the correct video to match the sentence they heard. Since all sentences were disambiguated by case-marking, this

measure allowed us to assess how well learners acquired the grammar of the novel language. Recruitment continued until the number of participants who achieved 70% accuracy on sentence comprehension tests on the final day of training reached 20 in each experiment. Participants who failed to pass this accuracy requirement (3 participants in Experiment 1 and 2 participants in Experiment 2) were removed from all analyses. The main pattern of results reported below did not change when participants who failed to pass the 70% criterion were included in the analysis.

Overall, learners achieved a high level of accuracy on the novel languages on the final day of training. Their mean comprehension accuracy was 97% in both experiments. Production performance showed a similarly high degree of accuracy, suggesting that the task was feasible. In Experiment 1, participants made 8.2% lexical mistakes and 3.5% grammatical mistakes on the final day of training. In Experiment 2, the rate of lexical mistakes was 12% and the rate of grammatical mistakes was 2.7%. The results reported below hold after removing all productions with lexical mistakes from the analysis.

Given that the languages in both experiments were acquired with a high degree of accuracy, any observed word order preferences are unlikely to be due to mistakes stemming from insufficient knowledge of the novel language.

### 4.2. Experiment 1: Production preferences in a verb-final miniature language

To assess learners' preferences in length-based ordering, we conducted a mixed effects regression analysis. We predicted learners' SO word order use from constituent length (all constituents short vs. object long, subject long vs. all other cases, Helmert coded), day of training (2 vs. 1, 3 vs. all other cases, Helmert coded), and their interactions. The model contained the maximal random effects structure justified by the data based on model comparison (by-subject random intercept, by-subject random slopes of day and constituent length). The same results were obtained when the maximal still converging model was used. Fixed effect correlations greater than 0.4 were observed between several predictors in the model. The results reported here did not change after the stepwise model reduction was used to reduce collinearity.

As expected under our hypothesis, participants' word order preferences in production revealed a bias for shorter dependencies. Despite receiving an unbiased input and having the opposite short-before-long preference in their native language, learners reduced the overall dependency length in the newly acquired language by introducing long-before-short ordering in their own productions (see Fig. 4). Across all three days of training, learners were significantly more likely to use SO order for sentences with long subject and short object phrases compared to other sentences types ($\widehat{ß}$=1.36, z=5.56, p<0.0001). Similarly, learners were significantly more likely to use SO order for sentences, in which both subject and object phrases were short, compared to sentences with short subject and long object phrases ($\widehat{ß}$=0.66, z=2.54, p<0.05). There was no main effect of day of

training (ps > 0.4) but day of training interacted with the effects of constituent length. On day 2, the difference in SO use between utterances with long subjects and utterances with no modification was significantly smaller as compared to day 1 ($\widehat{ß}$=-0.56, z=-4.02, p<0.0001). The difference in SO use for sentences with no modification as compared to the sentences with long objects was significantly larger on day 3 compared to day 2 ($\widehat{ß}$=0.17, z=2.12, p<0.05).

Simple effects tests revealed that learners used significantly more SO word order in utterances with long subjects compared to all other cases on all days of training (day 1: $\widehat{ß}$=1.96, z=5.66, p<0.0001; day 2: $\widehat{ß}$=0.84, z=3.56, p<0.0001; day 3: $\widehat{ß}$=1.27, z=5.21, p<0.0001). The difference in SO word order use for sentences with long objects compared to sentences with no modification reached significance only on the final day of training after participants became sufficiently fluent in the novel language (day 3: $\widehat{ß}$=1.02, z=3.36, p<0.0001).

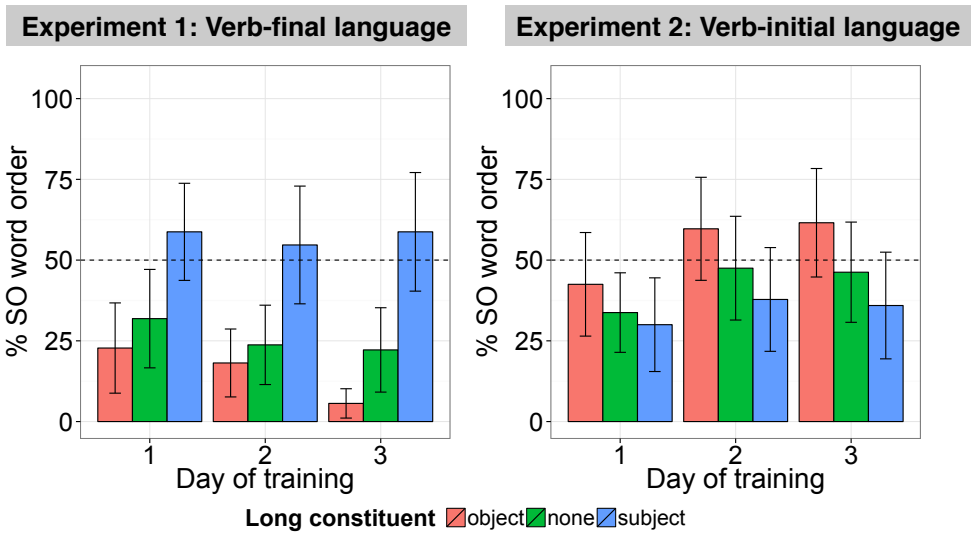


Figure 4: Subject-Object (SO) word order use in production in Experiment 1 (left panel) and Experiment 2 (right panel). The dotted line indicates the input proportion of SO order (equal across all sentence types and experiments). The error bars represent 95% confidence intervals.

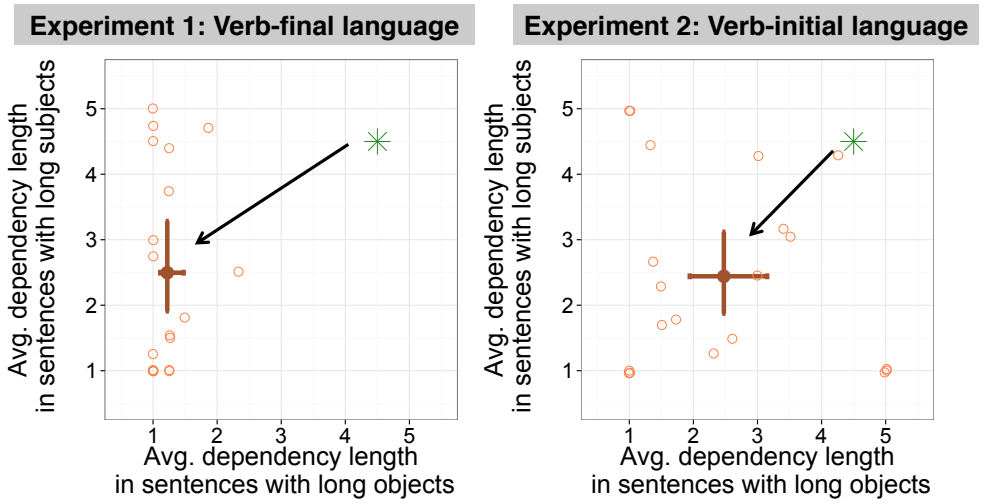

Figure 5: Average dependency length between the verb and its dependents in production (sentences with long constituents only) on the final (3rd) day of training. The green stars represent the expected average dependency length based on input statistics. Solid circles represent mean output languages produced by participants. Open circles represent output languages produced by individual participants. Error bars represent bootstrapped 95% confidence intervals.

### 4.3. Experiment 2: Production preferences in a verb-initial miniature language

Experiment 2 followed the same statistical procedure and variable coding as Experiment 1.

As expected under the dependency length minimization account, learners reduced the overall dependency length in the verb-initial language by introducing a short-before-long ordering preference in their productions—the opposite preference of that observed in the verb-final language in Experiment 1 (see Fig. 4). Across all days, learners were significantly less likely to use SO word order in sentences with long subject phrases and short object phrases compared to all other sentence types ($\widehat{ß}$=-0.44, z=-2.5, p<0.05). Similarly, learners were significantly less likely to use SO order in sentences with short subject and object phrases compared to sentences with long object and short subjects phrases ($\widehat{ß}$=-0.47, z=-2.01, p<0.05). This preference did not interact with day of training (ps > 0.2); nor there was a main effect of day of training (ps > 0.6).

Simple effects tests showed that the dispreference for SO word order in sentences with long subjects compared to all other sentence types was significant on day 2 ($\widehat{ß}$=-0.42, z=-2.27, p<0.05) and day 3 ($\widehat{ß}$=-0.52, z=-2.9, p<0.05) and marginally significant on day 1 ($\widehat{ß}$=-0.37, z=-1.76, p=0.08). The difference in SO use for sentences with no modification compared to sentences with long objects became significant with sufficient proficiency in the novel language -- on the final day of training ($\widehat{ß}$=-0.6, z=-2.47, p<0.05).

Thus, learners preferred opposite length-based constituent orders for verb-initial and verb-final languages. As Fig. 5 shows, these respective ordering preferences resulted in a substantial reduction of the average dependency lengths in the output languages produced by learners compared to the input.

## 5. Discussion

It has been hypothesized that the distribution of word orders in natural languages is shaped by biases for shorter grammatical dependencies (Hawkins, 2014). This hypothesis has been recently supported by evidence from large-scale cross-linguistic studies showing that languages exhibit properties consistent with this claim (Ferrer i Cancho, 2004; Futrell et al., 2015; Gildea & Temperley, 2010; Hawkins, 2014; Liu, 2008). The current study presents the first direct evidence for the hypothesized *causal link* between processing biases and patterns in cross-linguistic diversity. Learners in our experiments shared the same language background and received input languages with the same statistics but had different word order preferences depending on the verb (head) position

in the language. As predicted by the dependency length minimization hypothesis, learners preferred short-before-long ordering in the verb-initial language and long-before-short ordering in the verb-final language, which resulted in shorter dependencies in the two languages. This lends credibility to the hypothesis that the preference for short grammatical dependencies observed cross-linguistically originates in constraints on human information processing.

Prior work in linguistic typology has documented asymmetries in cross-linguistic word order preferences similar to those observed in our experiments: verb-initial languages like English and German tend to have an overall short-before-long preference, while verb-final languages like Japanese typically have a long-before-short preference (Hawkins, 2014). One caveat to these studies is that the observed preferences in adult speakers might be manifestations of grammaticalized structural preferences specifically acquired for this language, rather than evidence for biases in human information processing. Additionally, cross-linguistic correlations need to be interpreted with caution as many of them do not appropriately account for genetic or geographic dependencies between languages (Dunn, Greenhill, Levinson, & Gray, 2011; but see Croft, Bhattacharya, Kleinschmidt, Smith, & Jaeger, 2011 for a critique).

Learners' preferences in our experiments seem to be driven by an underlying preference for short grammatical dependencies. This preference introduces a small shift into learners' productions, thus providing a seed for the cross-linguistically observed preference for dependency length minimization. An important open question for future

research is whether these changes indeed accumulate as the language is transmitted over generations of speakers (as assumed here), thereby causing gradual language change over historical time (cf. Christiansen & Chater, 2008; Kirby et al., 2015).

Can our findings be accounted for by learners' native language preferences? Native language transfer effects have been widely attested in second language acquisition (Schepens, Van der Silk, & Van Hout, 2015; for a review see Pajak, Fine, Kleinschmidt, & Jaeger, in press) and thus present a serious consideration when interpreting our results. The native language of our participants (English) has an overall short-before-long ordering preference. This could thus explain the result of the verb-initial miniature language in Experiment 2, but not the inverse long-before-short preference in the verb-final miniature language in Experiment 1. This rules out direct surface-based transfer of overall word order preferences from English to the miniature languages in our experiments as a source of the observed effects (see SI for more details). An abstract principle-based transfer from the native language is, however, possible. The overall length of grammatical dependencies in English is close to the theoretical *minimum* (Gildea & Jaeger, under review; Gildea & Temperley, 2010). Thus, it is possible that native speakers of English are especially attuned to the principle of dependency length minimization and are readily extending this abstract preference to the novel miniature languages. Future work addressing this possibility is needed and would inevitably involve cross-linguistic data from speakers of languages that express a preference for shorter dependencies less strongly (e.g., German or Japanese).

Learners' preferences in our experiments closely mirror the cross-linguistic asymmetry in length-based ordering based on the verb (head) position. Our results, thus add to a growing body of literature suggesting that considerations of processing and communicative efficiency influence the way languages change over time and shape the synchronic distribution of cross-linguistic patterns (Fedzechkina et al., 2012; Fedzechkina et al., 2016; Futrell et al., 2015; Hawkins, 2014; Kemp & Regier, 2012; Kirby et al., 2015; Piantadosi, Tily, & Gibson, 2011).

**Author contributions**. M. Fedzechkina and T. Florian Jaeger developed the study concept and contributed to the design. B. Chu conducted the study and collected the data. M. Fedzechkina analyzed the data. All authors contributed to writing the manuscript and approved the final version of the manuscript for submission.

**Acknowledgements.** We thank Madeline Clark and Emily Eck for their help coding subject data; Ivy Gregory for his help with audio stimuli; Irene Minkina, Andy Wood, and Cassandra Donatelli for their help with video stimuli.

**Funding statement**. This work was supported in part by NSF CAREER grant IIS-1150028 to T. Florian Jaeger.

**Declaration of conflicting interests**. The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

**References:**


Arnold, J. E., Wasow, T., Losongco, T., & Ginstrom, R. (2000). Heaviness vs. Newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*(1), 28-55.

Bartek, B., Smith, M., Lewis, R., & Vasishth, S. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1178-1198.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language Acquisition: the State of the Art* (pp. 173-218). Cambridge: Cambridge University Press.

Blake, B. J. (2001). *Case*. Cambridge: CUP.

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 119): The MIT press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behav Brain Sci, 31*(05), 489-509. doi:10.1017/S0140525X08004998

Croft, W., Bhattacharya, T., Kleinschmidt, D., Smith, D. E., & Jaeger, T. F. (2011). Greenbergian universals, diachrony and statistical analyses. *Linguistic Typology, 15*(2), 433-453. doi:10.1515/lity.2011.029

Culbertson, J., & Adger, D. (2014). Language learners privilige structured meaning over surface frequency. *Proc Natl Acad Sci USA, 111*(16), 5842-5847.

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

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*(2), 193-210.

Dryer, M., & Haspelmath, M. (Eds.). (2011). *The world atlas of language structures online*. Munich: Max Planck Digital Library.

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*(7345), 79-82. doi:10.1038/nature09923

Fedzechkina, M., Jaeger, T., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proc Natl Acad Sci USA, 109*(44), 17897-17902. doi:10.1073/pnas.1215776109

Fedzechkina, M., Newport, E., & Jaeger, T. (2016). Balancing effort and information transmission during language acquisition: Evidence from word order and case-marking. *Cognitive Science*. doi:doi: 10.1111/cogs.12346

Ferrer i Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E, 70*, 056135.

Fodor, J. D. (2001). Setting syntactic parameters. *The Handbook of contemporary syntactic theory*, 730-767.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proc Natl Acad Sci USA, 112*(33), 10336-10341.

Gildea, D., & Jaeger, T. F. (under review). Language structure shaped by the brain: Human languages order information efficiently. *Proceedings of the Royal Society B*.

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Sci, 34*(2), 286-310.

Givón, T. (1991). Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Stud Lang, 15*(2), 335-370. doi:10.1075/sl.15.2.05giv

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of human language* (pp. 73-113): MIT Press.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science, 29*(2), 261–290.

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

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

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science, 336*(6084), 1049–1054.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87-102.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Cognitive Science, 9*, 159-191.

Lohse, B., Hawkins, J., & Wasow, T. (2004). Processing domains in English verb-particle construction. *Language, 80*(2), 238-261.

Pajak, B., Fine, A. B., Kleinschmidt, D., & Jaeger, T. (in press). Learning additional languages as hierarchical inference: insigts from L1 processing. *Language Learning*.

Pajak, B., & Levy, R. (2014). The role of abstraction in non-native speech perception. *Journal of Phonetics, 46*, 147-160.

Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA, 108*(9), 3526. doi:10.1073/pnas.1012551108

Ros, I., Santesteban, M., Fukumura, K., & Laka, I. (2015). Aiming at shorter dependencies: The role of agreement morphology. *Language, Cognition, and Neuroscience, 30*(9), 1156-1174.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Schepens, J., Van der Silk, F., & Van Hout, R. (2015). L1 and L2 distance effects in learning L3 Dutch. *Language Learning*, 1-33.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart, Winston.

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

Wasow, T. (2002). *Postverbal Behavior*. Stanford: CSLI Publications.

Yamashita, H., & Chang, F. (2001). Long before short preference in the production of a head-final language. *Cognition, 81*(2), B45-B55.