Human information processing as cause of language change: Learners prefer shorter dependencies

Maryia Fedzechkina*,[1], Becky Chu[2], and 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

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Languages across the world exhibit both striking diversity and abstract commonalities. Whether and how these commonalities are shaped by potentially universal principles of the human information processing has been of central interest in the biological, linguistic, and psychological sciences. Recent research has identified one abstract property that applies to all languages studied so far: although human languages differ greatly in how they order words in sentences, these superficially different orders seem to result in short grammatical dependencies between words. As short dependencies make sentences easier to process, these findings raise the possibility that languages are shaped by a deeper organizational principle rooted in human information processing. The current study takes the next critical step and tests this causal link. We ask whether a preference for shorter dependencies operates during language acquisition, causing learners to deviate from the input they receive. We find that learners exposed to novel miniature languages systematically restructure these languages to reduce the lengths of grammatical dependencies. The innovations introduced by learners thus provide direct evidence that processing preferences can shape language change and serve as the causal origin of cross-linguistic correlational patterns.

language universals | language processing | learning biases | artificial language learning

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 origins of these similarities has been the central question in the biological and language sciences (1-4). One proposal holds that cross-linguistic distributions reflect processing constraints and biases (5-7). A domain for which this proposal has been formalized and made testable is sentence word order. The order of words in a sentence differs across languages, but seems to exhibit certain generalizations. Specifically, there is now evidence that languages order words in such a way that the *grammatical dependencies* between the words are kept short (5, 8-11).

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 ('boy') and one with the direct object ('ball'). Efficient processing of grammatical dependencies is central to the human ability to communicate complex thoughts, as they determine how the information encoded in words is to be combined into the overall meaning of a sentence. It is now known that dependency length (i.e., the distance between the head and its dependent) affects processing efficiency: longer dependencies are associated with greater processing difficulty than shorter dependencies (12, 13), presumably due to increased memory demands (14, 15). 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 (16, 17), while speakers of verb-final languages like Japanese typically have a long-before-short preference (18-20). This verb-dependent ordering reduces the average dependency length in a sentence (see Fig. 1).

Crucially, recent large-scale computational studies have shown that this preference is not limited to production preferences under meaning equivalence. Rather, the *grammars* of human languages 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 (8-11, 21), with some languages being close to the theoretical *minimum* of dependency length (8, 22).

Taken together, these findings provide evidence for a *correlation* between memory demands during incremental language processing and cross-linguistic grammatical constraints on word order. This correlation raises a possibility that natural languages across the world are shaped by a deep organizational principle grounded in human information processing. The present study takes the next critical step in identifying the hypothesized *causality* underlying this correlation. Two experiments test whether the cross-linguistic preference for 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

Significance

Uncovering factors that make human language uniquely distinct from other animal communicative systems has been a fundamental question in language sciences. Universal principles of human information processing could be one such factor. Recent research has identified correlational patterns between processing ease and a preference for shorter dependencies in all languages studied so far. The present study takes the next crucial step and provides direct evidence that a preference for shorter dependencies operates during language acquisition. For the first time, we begin to see a complete chain, from a likely universal cognitive principle known to affect language processing, to this preference biasing learners to deviate from the input they receive, to these deviations causing cross-linguistic patterns over historical time.

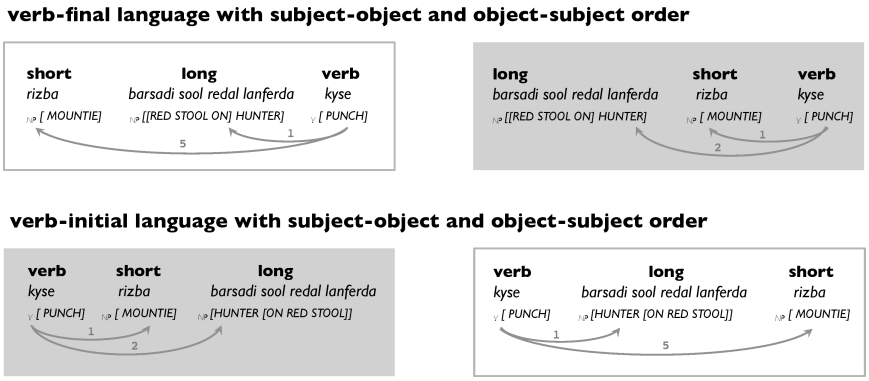

Fig. 1. Comparison of dependency lengths for two possible argument orderings (subject, object) in verb-final vs. verb-initial languages. All sentences express the same meaning. Arches represent grammatical dependencies between the verb and the head of its two arguments. Numbers represent dependency lengths, 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.

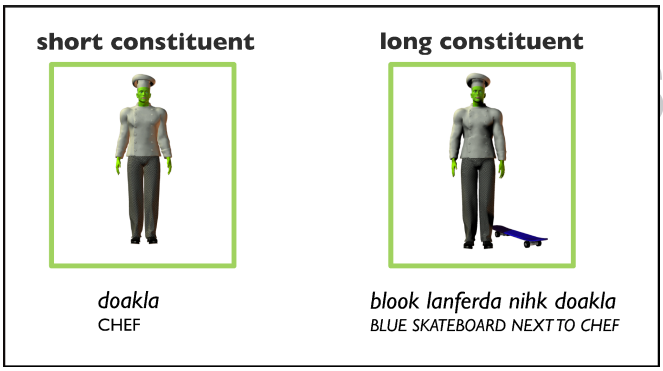

Fig. 2. Illustration of constituent length manipulation in the experiments. The visual referent on the left can be described with a short phrase. The referent on the right requires a complex long phrase. Example descriptions are shown for the verb-final miniature language from Experiment 1 (provided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

the original input. This would introduce small deviations towards a grammar with shorter dependencies into the input for the next generation of learners. 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 for 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. 6).

We test the first step in this chain. We use the miniature artificial language learning paradigm (23, 24) 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 (25, 26). 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 (23, 24, 27-32). 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.

Results

In two experiments, monolingual native speakers of English learned a miniature artificial language by watching short videos describing simple transitive actions performed by two human actors (e.g., 'chef punch referee') and hearing their descriptions in the novel miniature language (see Materials & Methods and SI). The miniature language was either verb-final (Experiment 1) or a verb-initial (Experiment 2). Both languages had flexible word order, so that subject-object (SO) and object-subject (OS) orders occurred equally frequently in the input. In both languages, half of the sentences in the input had long subject and object phrases (long-long sentences), and half of the sentences had short subject and object phrases (short-short sentences). 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 word orders (SO and OS) occurred equally often in short-short and long-long sentences. Further details about the input languages are given in the Materials and Methods section and SI.

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 (16, 17, 33, 34). Furthermore, any observed word order preferences are unlikely to be due to mistakes stemming from insufficient knowledge of the novel language as the languages in the two experiments were learned with a high degree of accuracy (mean comprehension accuracy on the final day of training was 97%). 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.

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

Participants' word order preferences in production supported our hypothesis. Despite receiving an unbiased input and having the opposite short-before-long bias in their native language,

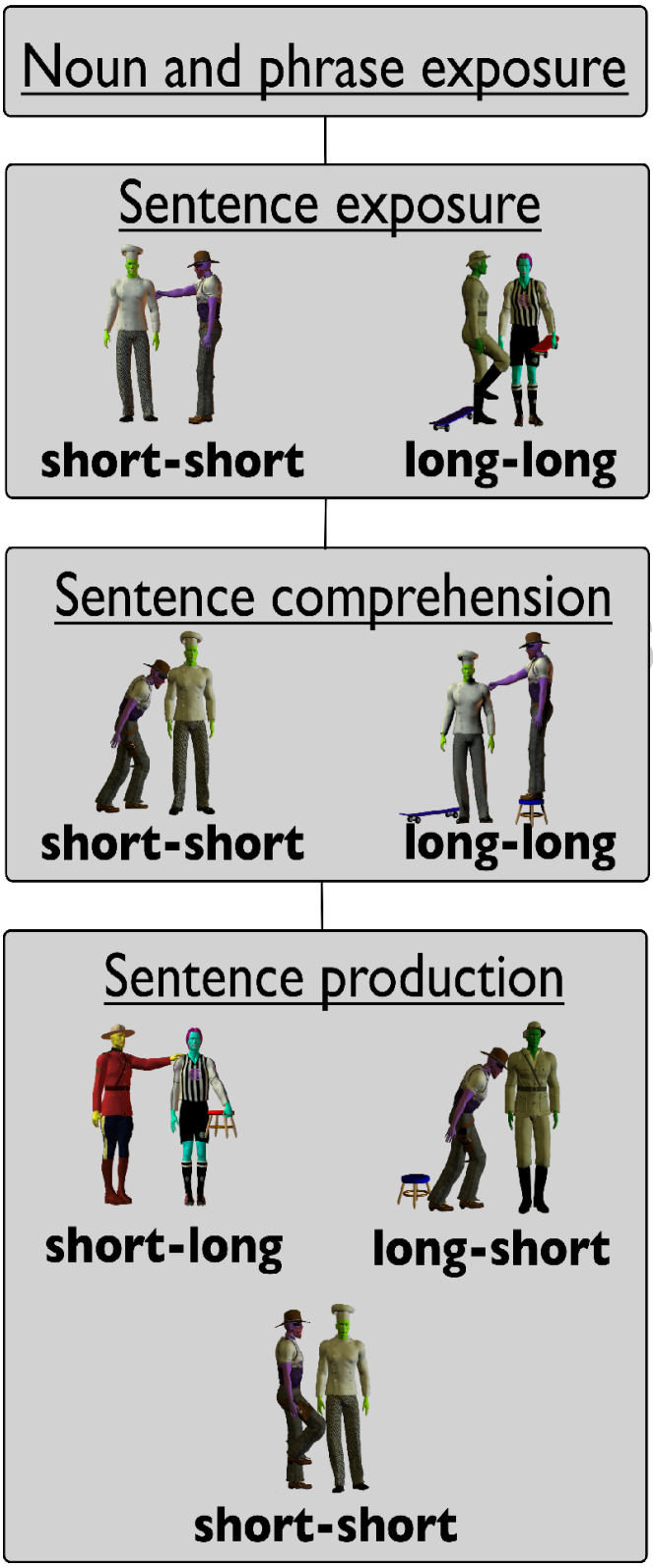

Fig. 3. Each participant was exposed to a novel miniature language following the procedure shown here on each of three separate visits to the lab (images represent still pictures of video stimuli used in the experiments). The measure of interest was the word orders learners produced in the Sentence Production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

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 ($\hat{\beta}$=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 ($\hat{\beta}$=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 length (see SI).

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

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 ($\hat{\beta}$=-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 ($\hat{\beta}$=-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).

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.

Discussion

It has been hypothesized that the distribution of word orders in natural languages is shaped by biases for shorter grammatical dependencies (5, 35). There is now evidence from large-scale cross-linguistic studies that languages exhibit properties consistent with this claim (5, 8-11, 21, 35). The present 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 (5, 35). 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 (36; but see (37) for a critique).

Learners' preferences in our experiments seem to be driven by an underlying preference for short grammatical dependencies.

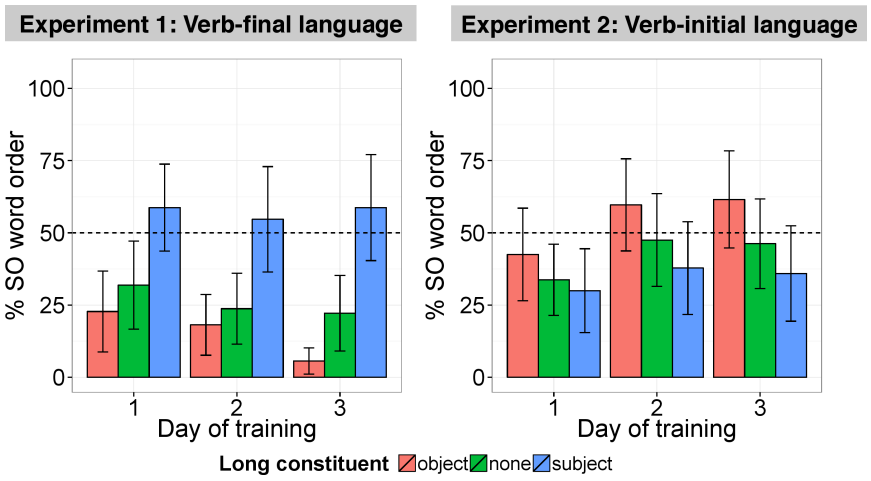

Fig. 4. Subject-Object (SO) word order use in production in Experiment 1 (left panel) and Experiment 2 (right panel). The dotted line indicates the input proportion of SO order (equal across all sentence types and experiments). The error bars represent 95% confidence intervals.

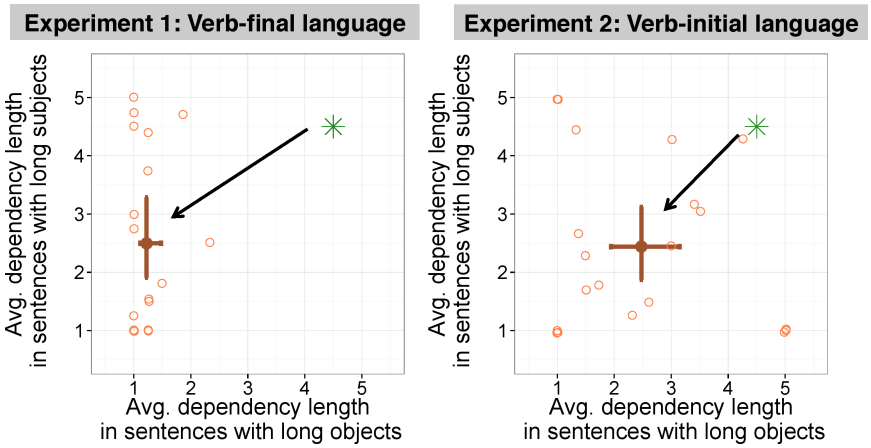

Fig. 5. Average dependency length between the verb and its dependents in production (sentences with long constituents only) on the final (3rd) day of training. The green stars represent the expected average dependency length based on input statistics. Solid circles represent mean output languages produced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

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. 6, 7, 24, 38).

Can our findings be accounted for by learners' native language preferences? Native language transfer effects have been widely attested in second language acquisition (39; for a review see 40-41). 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* (8, 22). 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 (5, 9, 28, 30, 31, 35, 42-45).

Materials and Methods

Participants

Participants in both 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 (31), recruitment continued until 20 participants, who successfully learned the language, were enrolled in each experiment (see SI for details). Recruiting and execution of the studies was approved by the Research Subjects Review Board at the University of Rochester.

Procedure

All 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.

Input Languages. Languages in both experiments 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 (see SI for more details). Like many languages with flexible word order (46), 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. Both languages contained adpositional phrases (see Fig. 2, e.g., 'chef next-to blue skateboard'). The order of the adposition (e.g., 'next-to') relative to its dependent ('blue skateboard') and head ('chef') followed cross-linguistically common patterns (47), as shown in Figure 1. The verb-final language (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 (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 modification; 50% of training scenes) or two 'long' constituents (i.e., subject and object with modification; 50% of training scenes). 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).

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

1. Newport E (2011) The modularity issue in language acquisition: A rapprochement? Comments on Gallistel and Chomsky. *Language Learning and Development* 7:279-286.
2. Chomsky N (1965) *Aspects of the Theory of Syntax* (The MIT press).
3. Greenberg J (1963) Some universals of grammar with particular reference to the order of meaningful elements. *Universals of human language*, ed Greenberg J (MIT Press), pp 73-113.
4. Chomsky N (2011) Language and other cognitive systems. What is special about language? *Language Learning and Development* 7:263-278.
5. Hawkins JA (2014) *Cross-linguistic variation and efficiency* (Oxford University Press, Oxford).
6. Christiansen MH & Chater N (2008) Language as shaped by the brain. *Behav Brain Sci* 31(05):489-509.
7. Kirby S (1999) *Function, selection, and innateness: The emergence of language universals* (Oxford University Press, USA).
8. Gildea D & Temperley D (2010) Do grammars minimize dependency length? *Cognitive Sci* 34(2):286-310.
9. 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.
10. Liu H (2008) Dependency distance as a metric of language comprehension difficulty. *Cognitive Science* 9:159-191.
11. Ferrer i Cancho R (2004) Euclidean distance between syntactically linked words. *Physical Review E* 70:056135.
12. Grodner D & Gibson E (2005) Consequences of the serial nature of linguistic input. *Cognitive Science* 29(2):261–290.
13. Demberg V & Keller F (2008) Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2):193-210.
14. 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.
15. Lewis RL, Vasishth S, & Van Dyke JA (2006) Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10(10).
16. Arnold JE, 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.
17. Wasow T (2002) *Postverbal Behavior* (CSLI Publications, Stanford).
18. Yamashita H & Chang F (2006) Sentence production in Japanese. *Handbook of East Asian Psycholinguistics (Volume 2, Japanese)*, eds Nakayama M, Mazuka R, & Shirai Y (Cambridge University Press, Cambridge, UK).
19. 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.
20. Choi H-W (2007) Length and Order: A Corpus Study of Korean Dative-Accusative Construction. *Discourse and Cognition* 14:207-227.
21. Gildea D & Temperley D (2007) Optimizing Grammars for Minimum Dependency Length. *Proceedings of the 45th Annual Conference of the Association for Computational Linguistics (ACL 2007)*, (The Association for Computational Linguistics), pp 184-191.
22. Gildea D & Jaeger TF (under review) Language structure shaped by the brain: Human languages order information efficiently. *Proceedings of the Royal Society B*.
23. Hudson Kam C & Newport E (2009) Getting it right by getting it wrong: When learners change languages. *Cognitive Psychol* 59(1):30-66.
24. Kirby S, Cornish H, & Smith K (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc Natl Acad Sci USA* 105(31):10681.
25. Saffran J, Aslin R, & Newport E (1996) Statistical learning by 8-month-old infants. *Science* 274(5294):1926-1928.
26. Pajak B & Levy R (2014) The role of abstraction in non-native speech perception. *Journal of Phonetics* 46:147-160.
27. Culbertson J, Smolensky P, & Legendre G (2012) Learning biases predict a word order universal. *Cognition* 122:306-329.
28. Kirby S, Tamariz M, Cornish H, & Smith K (2015) Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141:87-102.
29. Smith K & Wonnacott E (2010) Eliminating unpredictable variation through iterated learning. *Cognition* 116(3):444-449.
30. Fedzechkina M, Newport E, & Jaeger T (in press) Balancing effort and information transmission during language acquisition: Evidence from word order and case-marking. *Cognitive Science*.
31. 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.
32. Culbertson J & Adger D (2014) Language learners privilige structured meaning over surface frequency. *Proc Natl Acad Sci USA* 111(16):5842-5847.
33. Lohse B, Hawkins J, & Wasow T (2004) Processing domains in English verb-particle construction. *Language* 80(2):238-261.
34. Wasow T (1997) Remarks on grammatical weight. *Language Variation and Change* 9(1):81-105.
35. Hawkins JA (2004) *Efficiency and Complexity in Grammar* (Oxford University Press., Oxford).
36. Dunn M, Greenhill SJ, Levinson SC, & Gray RD (2011) Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345):79-82.
37. Croft W, Bhattacharya T, Kleinschmidt D, Smith DE, & Jaeger TF (2011) Greenbergian universals, diachrony and statistical analyses. *Linguistic Typology* 15(2):433-453.
38. Reali F & Griffiths T (2009) The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111(3):317-328.
39. Schepens J, Van der Silk F, & Van Hout R (2015) L1 and L2 distance effects in learning L3 Dutch. *Language Learning*:1-33.
40. Odin T (2013) Crosslinguistic influence in second language acquisition. *The encyclopedia of applied linguistics*, ed Chapelle C (Blackwell, Malden, MA), pp 1562–1568.
41. Pajak B, Fine AB, Kleinschmidt D, & Jaeger T (in press) Learning additional languages as hierarchical inference: insigts from L1 processing. *Language Learning*.
42. Tily H (2010) The role of processing complexity in word order variation and change. Ph.D. (Stanford University).
43. Kemp C & Regier T (2012) Kinship categories across languages reflect general communicative principles. *Science* 336(6084):1049–1054.
44. Slobin D (1977) Language change in childhood and in history. *Language Learning and Thought*, ed Macnamara J (Acad. Press, New York), pp 185-214.
45. Piantadosi S, Tily H, & Gibson E (2011) Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA* 108(9):3526.
46. Blake BJ (2001) *Case* (CUP, Cambridge).
47. Dryer M & Haspelmath M eds (2011) *The world atlas of language structures online* (Max Planck Digital Library, Munich).