

Language learners restructure their input to facilitate efficient communication

Maryia Fedzechkina^{a,1}, T. Florian Jaeger^{a,b,1}, and Elissa L. Newport^{a,c,1}

Departments of ^aBrain and Cognitive Sciences and ^bComputer Science, University of Rochester, Rochester, NY 14627; and ^cDepartment of Neurology, Georgetown University, Washington, DC 20007

Contributed by Elissa L. Newport, September 14, 2012 (sent for review August 1, 2012)

Languages of the world display many structural similarities. We test the hypothesis that some of these structural properties may arise from biases operating during language acquisition that shape languages over time. Specifically, we investigate whether language learners are biased toward linguistic systems that strike an efficient balance between robust information transfer, on the one hand, and effort or resource demands, on the other hand, thereby increasing the communicative utility of the acquired language. In two experiments, we expose learners to miniature artificial languages designed in such a way that they do not use their formal devices (case marking) efficiently to facilitate robust information transfer. We find that learners restructure such languages in ways that facilitate efficient information transfer compared with the input language. These systematic changes introduced by the learners follow typologically frequent patterns, supporting the hypothesis that some of the structural similarities found in natural languages are shaped by biases toward communicatively efficient linguistic systems.

language universals | learning biases | efficient information transmission | communicative pressures

One of the central objectives of modern linguistics is to identify the principles that characterize possible human languages. To this end, linguists have examined languages of the world to find patterns that recur across languages (“linguistic universals”). The origins of such recurring patterns have been the subject of long-standing debate in linguistics and cognitive science. One view holds that language universals arise because of innate constraints, specific to language and not characteristic of other aspects of cognition (1, 2). A second view argues that languages are shaped over time by constraints on human cognitive mechanisms and pressures associated with language acquisition and use. A variety of such cognitive pressures have been proposed to constrain the space of possible language structures, such as learnability (3–5), memory limitations (6), constraints on processing and perception (7, 8), and considerations of efficient communication (9–11). On this view, language structures that increase the learnability of a language, reduce its processing complexity, or ensure efficient communication are more likely to be observed cross-linguistically (for a review, see ref. 12).

Arguments for linguistic universals of either type (domain-specific or not) have primarily been based on generalizations across typological data (8, 13, 14). This approach has been challenged because of data sparsity (the number of thoroughly documented languages is relatively small) and statistical dependencies between the data points (genetically or geographically related languages tend to share linguistic features). Indeed, recent work has questioned whether there is any evidence for linguistic universals when these dependencies are appropriately accounted for (15). However, these approaches too have their shortcomings, including a lack of power (16, 17).

An alternative to the typological test of linguistic universals is offered by the artificial language learning paradigm (18–20). This paradigm, in which participants learn miniature languages in the laboratory, has provided evidence that certain grammatical systems are easier to learn than others, consistent with hypothesized linguistic universals (21–28). General learning biases (for example,

a bias toward more regular lexical or grammatical systems) can also be demonstrated in artificial language learning (18, 19, 29, 30).

Here, we focus on the hypothesis that one of the biases shaping languages over time is a preference for communicatively efficient linguistic systems. Successful communication occurs when the message intended by the speaker is correctly inferred by the listener. Although human communication takes place in the presence of both environmental and biological noise, robust information transfer can be achieved through redundancy in the linguistic signal (31). Furthermore, efficient information transfer implies more redundancy in a linguistic signal that encodes unexpected messages, compared with a linguistic signal that encodes more expected messages (32–35). The same inverse relation between expectedness and the amount or quality of the linguistic signal is also predicted if robust information transfer is balanced against the time and resources required for message encoding (11, 36). Indeed, languages around the world exhibit a variety of otherwise unexpected properties that facilitate efficient communication, as defined here (11, 32, 34, 37–42).

What is unknown, however, is how such properties enter the linguistic system. Two mutually compatible scenarios are possible (cf., refs. 9 and 34). Biases toward efficient communication operating throughout life can cause adults to subtly change the input they provide to the next generation. It is also possible that these biases toward efficient linguistic systems operate during language acquisition, leading learners to deviate slightly from the input they receive as they acquire the language. Despite the long history of these claims, strong tests of these two hypotheses are lacking.

Here, we address the latter possibility, that biases toward communicatively efficient linguistic systems operate during language acquisition. We investigate, in particular, whether learners use the formal devices of the input language in a way that increases the average rate of information transmission. More specifically, we ask whether learners alter the input language, providing additional cues to the intended meaning of the sentence when other properties of the sentence would likely cause processing difficulty or misinterpretation. Such deviations from the input language toward a linguistic system that makes more efficient use of redundancy in the linguistic signal could be a vehicle for language change over generations. If these changes shift the input language toward typologically common patterns, this would provide evidence that gradient linguistic universals can result from biases for efficient communication.

Differential Case Marking Systems

As a test case, we investigate the acquisition of differential case marking systems (43–45) found in a large number of natural

Author contributions: M.F., T.F.J., and E.L.N. designed research; M.F. performed research; T.F.J. and E.L.N. contributed new reagents/analytic tools; M.F. analyzed data; and M.F., T.F.J., and E.L.N. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. E-mail: mashaf@bcs.rochester.edu, fjaeger@bcs.rochester.edu, or eln10@georgetown.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1215776109/-DCSupplemental.

languages (e.g., Sinhalese, Spanish, Russian, and Hindi). Case marking is the addition of markers on nouns, typically prefixes or suffixes added to the noun stem, for example, to indicate which noun is the subject and which is the direct object of the verb. Differential case marking languages mark only certain types of subjects and direct objects, leaving others zero-marked. Although morphological case is, thus, optional in such systems, its occurrence and omission are highly principled and are generally associated with certain semantic properties of the referent such as animacy, definiteness, and person, as shown in Scheme 1:

Animacy scale: human > animate > inanimate

Definiteness scale: personal pronoun > proper name > other

Person scale: first, second > third

Referents that are higher on the dimensions in Scheme 1 are typically associated with the subject position, whereas referents that are lower on those dimensions typically occur as sentential objects. This mapping from scales of referential properties (e.g., human > animate > inanimate) to the grammatical function hierarchy (e.g., subject > object) is sometimes referred to as alignment. For atypical alignments, grammatical functions are more often signaled by case marking (43, 45, 46). Here, we investigate animacy effects. In differential case marking languages, inanimate subjects and animate objects (less typical alignments) are categorically case-marked, whereas animate subjects and inanimate objects (more typical alignments) are categorically not case-marked (43, 45, 46).

Optional case marking languages, such as Korean and Japanese, exhibit the same general tendency as differential case-marking languages but do so gradually. That is, subjects and objects are more or less likely to be case-marked depending on how typical their referents are for the grammatical function they carry (47).

These animacy effects in optional and differential case marking can be recast in terms of efficient information transfer through a noisy channel. Consider a simple transitive sentence, such as “The man the wall hit,” in a hypothetical language with flexible constituent order [a language in which subject–object–verb (SOV) and object–subject–verb (OSV) constituent orders are both permitted, e.g., German or Korean]. Here, the grammatical functions of “man” and “wall” cannot be identified based on the linear order of elements alone. If the intended message is that the man is hitting the wall, speakers can rely on listeners inferring the correct message because “the man” (animate) is a typical agent (the doer of an action), and “the wall” (inanimate) is a typical undergoer (the referent affected by an action). Case marking will add little to such a sentence. However, the less the relative animacy of referents itself biases listeners toward the intended message, the more important case marking becomes. This is most evident when animacy biases the listener toward the wrong interpretation (e.g., if the wall is hitting the man, for example, because it is falling onto the man). Similarly, case marking will help to facilitate successful communication when the noun referents rank equally on the animacy hierarchy (“The man the woman hit”). This logic extends to the cross-linguistically more typical case, in which constituent order provides some information (e.g., when subjects tend to precede objects): case marking can always be used to further reduce the uncertainty about the intended meaning, but its usefulness is highest if the other cues (e.g., constituent order, animacy) do not bias listeners toward the intended meaning. Thus, under our hypothesis, a referential expression should be more likely to receive overt case marking when its intended grammatical function is less expected, given other properties of the sentence including animacy (see also ref. 34).

We present two miniature artificial language learning experiments that test this hypothesis. Participants learned a miniature artificial language during four 45-min sessions, spread over

4 consecutive days, by watching short computer-generated videos and hearing their descriptions in the novel language (Fig. 1). This paradigm has been successfully used to investigate language acquisition in children (18, 19) and adults (22, 30, 29, 48). This work has found that adults typically match the statistics of the input miniature language very closely (18, 19). Here, however, we are interested in whether learners produce deviations from their input language when that language employs its formal devices in a way that leads to inefficient information transfer. Experiments 1 and 2 expose learners to languages that do not have efficient case marking. Will learners deviate from this input to make the language more communicatively efficient?

Specifically, we expose learners to languages that are inefficient versions of a verb-final language with flexible constituent order and optional case marking. The miniature languages used in our experiments resembled naturally occurring languages in that they had a dominant constituent order (SOV, 60% of all sentences) and a less frequent constituent order (OSV, 40% of all sentences). Like many verb-final languages with flexible constituent order, our miniature languages contained case marking. Crucially, however, our miniature languages deviated from naturally occurring languages in that case marking was not conditioned on animacy. In experiment 1, the grammatical object was optionally case-marked (in 60% of all input sentences). In experiment 2, the grammatical subject was optionally case-marked (also in 60% of the input). In

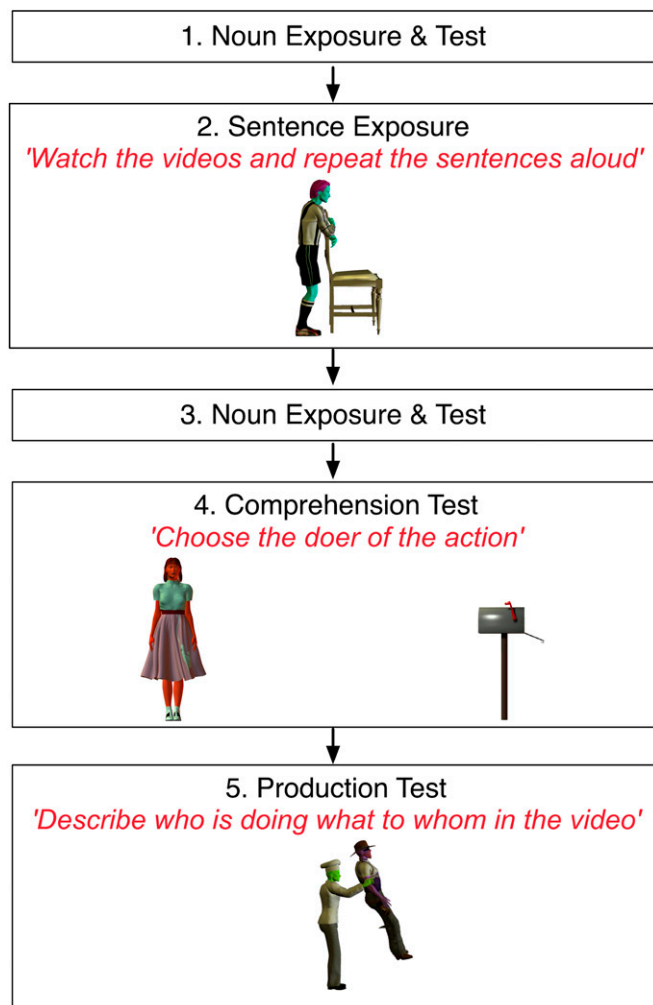


Fig. 1. Experimental procedure showing still images of the video stimuli used in the experiments.

both experiments, case marking appeared equally frequently on animate and inanimate noun phrases.

If learners are indeed biased to restructure the input language to increase its communicative efficiency, learners should introduce animacy-contingent case marking, that is, increase the use of case marking on referents that are less likely to carry the grammatical function intended by the speaker, while leaving more expected referent-to-grammatical function assignments zero-marked. Importantly, participants in our experiments were monolingual speakers of English. English has no productive case marking system. Although there are remnants of a former case marking system preserved in the pronominal system (e.g., he vs. him), English does not case-mark lexical nouns, such as those in our experiments. Crucially, English does not have optional case marking. So, if observed, the introduction of animacy-contingent case marking into the artificial language could not be attributable to transfer from the native language.

Experiment 1

All sentences represented simple transitive actions, such as “poke” or “hug,” performed by a human actor on either human or inanimate undergoers, which occurred equally often in the exposure (*Materials and Methods*). Because the language had flexible constituent order, sentences with human objects were ambiguous if the object was not case-marked, but sentences with inanimate objects could be disambiguated based on animacy even without a case marker.

If language users indeed try to communicate efficiently, they should restructure the language as they learn it, making it similar to differential object marking systems found in natural languages. In particular, if language learners are biased toward communicatively efficient linguistic systems, we would expect them to mark animate objects with an overt case marker more frequently than inanimate objects.

First, we examined data from the comprehension test, asking whether a primary function of case marking is to disambiguate the intended actor and undergoer. As expected, when both referents were animate and there was no case marking on the object, there were many misinterpretations of the intended meaning [53% mean accuracy, not significantly different from chance: $\chi^2(1) = 0.28$; $P = 0.59$; ns]. Overt object case marking significantly and substantially increased the accuracy of responses [88% accuracy, significantly different from chance: $\chi^2(1) = 63.82$; $P < 0.0001$].

We then examined the data from the production test to see what participants learned about the language (see *SI Text* for complete details on scoring and analysis). Do participants restructure the language in their productions to make more efficient use of its formal devices? Consistent with our hypothesis,

participants’ productions deviated from the input in that atypical objects were more likely to be case-marked (Fig. 2A). Learners used significantly more case markers on atypical (animate) objects than on typical (inanimate) objects across all days of testing ($\beta = 0.35$; $z = 2.27$; $P < 0.05$), even though this was not the pattern of their input language. This pattern of conditioning overt case marking on animacy closely mirrors the pattern commonly found in differential object marking systems (45, 47).

We also found that objects were more likely to be overtly case-marked if the constituent order was OSV ($\beta = 1.06$; $z = 2.14$; $P < 0.05$) (Fig. 2B). This pattern is opposite to the input distribution, where more case marking was used on objects in SOV sentences. There are several possible explanations of this result, some of which provide further support for our hypothesis. We postpone the discussion of this finding until after the discussion of experiment 2.

The observed effect of animacy is also compatible with an alternative explanation: the higher proportion of case marking on animate objects might arise because animate referents attract more visual attention (49), which might cause participants to learn case marking earlier or more successfully for animate referents. This concern was addressed in experiment 2, which explored optional subject case marking. If the results from experiment 1 are attributable to a bias to case-mark the atypical, as we hypothesize, then the opposite pattern should hold for optional subject case marking. We would expect participants to be more likely to use case markers on inanimate subjects, while leaving the typical animate subjects more frequently zero-marked. In contrast, if the observed behavior is attributable to increased attention to animate referents, we would expect participants to case-mark animate referents more frequently in both experiments.

Experiment 2

The input language in experiment 2 was the complement of the language used in experiment 1. In experiment 2, the animacy of subject varied (50% of subjects were animate and 50% were inanimate); objects were always inanimate. Sentential subjects were optionally case-marked independently of animacy, whereas objects were always zero-marked. All other aspects of the input grammar were the same as in experiment 1.

We first analyzed data from the comprehension test, asking about listeners’ accuracy in decoding the intended meaning. As in experiment 1, learners showed chance performance [52% mean accuracy; $\chi^2(1) = 0.71$; ns] when the referents were matched for animacy and the subject was not overtly case-marked. Performance was substantially improved and was significantly above chance when subjects were case-marked [94% mean accuracy;

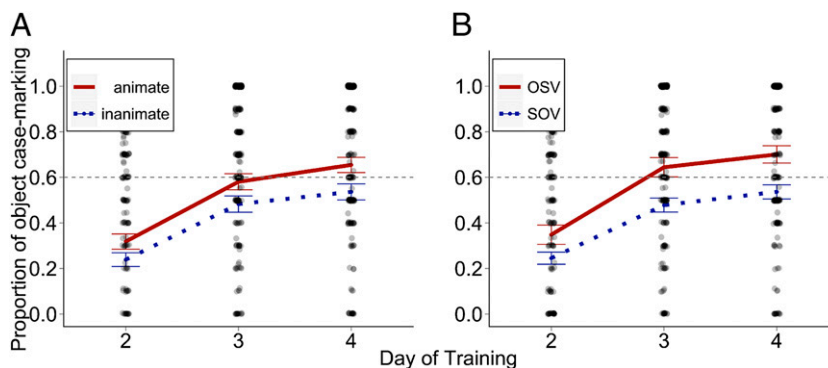


Fig. 2. Overt case marking by animacy of object (A) and constituent order (B) in production in experiment 1. Lines represent condition means, and dots represent overall subject means. Error bars represent 95% confidence intervals. The dashed line indicates proportion of case marking provided in the input (invariant across animacy).

$\chi^2(1) = 91.7; P < 0.0001$] or were animate [82% mean accuracy; $\chi^2(1) = 35.8; P < 0.0001$].

Fig. 3 shows the data from participants' productions (see *SI Text* for complete details on scoring and analysis). On the first day of testing, animate referents were case-marked significantly more frequently than inanimate referents ($\beta = -0.33; z = -2.5; P < 0.05$). This behavior is consistent with the alternative hypothesis, that the higher proportion of case marker use with animate referents may be driven by properties associated with animacy. However, this bias to case-mark animate referents, evident at early stages of learning, gradually weakens as training continues, giving way to a bias toward efficient information transfer, which emerges through language exposure as learners become more proficient. This is evidenced by a significant day \times animacy interaction ($\beta = 0.22; z = 2.87; P < 0.01$): as expected under our hypothesis, on the final day of training learners show the opposite preference and use more case marking on atypical inanimate subjects than on animate subjects.

We also examined case marker use in relationship to constituent order (Fig. 3B). In experiment 1, we observed more frequent object case marking in the OSV order. Such word-order contingent case marking could be driven by at least two biases. First, as hypothesized above, OSV order may bias the listener to an incorrect grammatical function assignment; hence, case marking is used to avoid potential miscommunication. Alternatively, word-order contingent case marking may reflect a bias to mention disambiguating information as early as possible in the sentence (8). For experiment 2, such a bias toward early disambiguation makes the opposite prediction compared with experiment 1: in a language with subject case marking, more frequent case marking should be observed in the SOV order because this provides information about grammatical function assignment earlier in the sentence.

The results of experiment 2 suggest that both biases are at play. There was a main effect of word order: overall, significantly more subjects were overtly marked when the constituent order was SOV ($\beta = 0.93; z = 2.40; P < 0.05$), which is indicative of a bias to provide disambiguating information at the earliest possible moment. This bias, however, gradually weakened as training continued, as suggested by the significant word order \times day interaction ($\beta = -0.16; z = -4.11; P < 0.001$). There was no significant preference to differentially case-mark subject referents depending on sentence word order on the final day of training ($\beta = 0.60; z = 1.53; P = 0.13; ns$). This might indicate a point at which participants' productions would start to reflect the bias to mark the atypical if training continued.

Importantly, the more complex (one might say, weaker) results of experiment 2 actually parallel quite nicely the typological data

from natural languages. Differential object marking is cross-linguistically highly consistent: languages with animacy-contingent differential object case marking tend to follow the pattern found in experiment 1 (50). In contrast, differential subject marking in natural languages is typologically less clear-cut, and this was also true of learners in experiment 2. The two competing acquisition biases observed in experiment 2 (a bias to case-mark animate referents and a bias to case-mark less expected referent-to-grammatical function assignments) are manifested typologically as well. Many languages, such as Mangarayi (51), overtly mark inanimate subjects and leave animate subjects zero-marked, but there are languages (e.g., Samoan) that have been claimed to show the opposite pattern (52).

Discussion

It has long been hypothesized that communicative pressures on language can operate during acquisition (9). The studies presented here provide experimental evidence supporting and clarifying this hypothesis. Our results suggest that language learners are biased toward communicatively efficient linguistic systems and restructure the input language in a way that facilitates information transfer, in line with recent information-theoretic approaches to language production (33–35). In our experiments, this bias affects the acquisition of an optional case marking system: although case marking in the input language is independent of animacy, learners showed a tendency to condition case marking on animacy, with the less expected alignments of animacy and grammatical function (inanimate subjects or animate objects) becoming more likely to be case-marked. Note that learners could instead have generalized case marking to all nouns, regardless of animacy. This would have maximized the chance of communicative success at the expense of effort and, possibly, processing speed (reducing the rate of information transmission), because case markers would be produced even when the intended meaning could be inferred in their absence. However, very few participants showed full case marker generalization (*Materials and Methods*), suggesting that the tradeoff between successful communication and effort was indeed at work during learning. The observed bias toward efficient linguistic systems is not reducible to previously documented tendencies of learners to regularize inconsistent structures (18, 19), biases to reduce the representational complexity of linguistic systems (29, 30), or a native language bias because we exposed native speakers of English (a language with no case marking on nouns) to an artificial language with optional case marking.

Our results also bear on the discussion of ambiguity avoidance in sentence production. Previous work suggests that speakers do not avoid ambiguity that is rapidly resolved through contextual information and world knowledge (53, 54). At the same time,

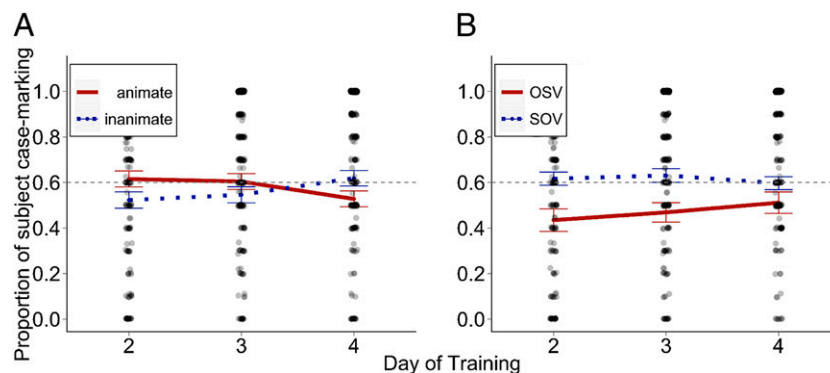


Fig. 3. Overt case marking in production by animacy of subject (A) and constituent order (B) in experiment 2. Lines represent condition means, and dots represent overall subject means. Error bars represent 95% confidence intervals. The dashed line indicates proportion of case marking provided in the input (invariant across animacy).

languages seem to avoid alternations that cause systemic ambiguity (55, 56). This has led some to hypothesize that ambiguity avoidance emerges during language acquisition (57). Our results support this view: learners avoided ambiguity that would have remained globally unresolved.

Our experiments raise questions about the precise nature of the mechanism underlying the biases we observed. One possibility is that the language production system is organized (either innately or through learning) to prefer efficient information transfer (32, 34, 35, 58, 59). Another possibility is that learners misinterpreted some of the sentences they were exposed to, altering the characteristics of the input from which they learned. In accord with the comprehension data, misinterpretations would have been most common in the absence of case marking and, in particular, when the animacy of the two arguments did not bias learners toward the intended message (i.e., in the less typical alignments). This would lead to higher (perceived) proportions of case marking for each type of atypical arguments (e.g., each animate object noun in experiment 1) compared with typical arguments. However, in our experiments the meaning of the sentence was always represented by an accompanying video, thereby unambiguously conveying the intended meaning. Given that there also was no time pressure, it is relatively unlikely that misinterpretations were sufficiently frequent to create the observed effect. More probable, however, is that this type of “misinterpretation” arises later, when form-meaning mappings are reconstructed from memory (60). Further work is necessary to distinguish between these and other mechanisms to explain our results.

Regardless of what mechanisms underlie the bias toward efficient languages in our experiments, our results suggest that learners do introduce typologically common patterns into the language. The learning outcomes in our experiments closely mirror natural phenomena, such as optional case marking systems found in Japanese and Korean, where animate objects and inanimate subjects are more likely to receive overt case marking (47). The close correspondence between the patterns observed during acquisition and those found in typological data suggests that some of the properties of natural languages may be shaped by learning biases that stem from a preference for communicatively efficient linguistic systems.

In this way, our results complement previous artificial language learning studies of phonology (23, 27), lexical, and syntax acquisition (22, 26, 29, 30) showing behavioral evidence for linguistic universals. Together, these and our studies demonstrate the power of the artificial language learning paradigm as a complement to typological work on linguistic universals (cf., ref. 17). The biases we have observed during the acquisition of optional case marking provide a possible mechanism for patterns observed cross-linguistically (37, 40, 41) and during native language production by adult speakers (32–35, 55, 61).

Materials and Methods

Experiments 1 and 2 used identical procedures. They differed only in certain aspects of the input languages presented to participants.

Participants. Participants in experiments 1 and 2 were undergraduate students at the University of Rochester, all of whom were monolingual native speakers of English. Informed consent was obtained from all participants. Recruiting and execution of the study were approved by the Research Subjects Review Board of the University of Rochester. Each participant was tested in only one of the two experiments. Participants were paid \$5 on days 1–3 of the experiment and \$25

upon completion of the fourth and final session. Twenty-nine participants completed experiment 1, with one participant excluded because of experimenter error, three participants excluded for failing to achieve a 70% comprehension accuracy requirement (suggesting that overall they had not learned the language sufficiently), and five participants excluded for using the case marker in all or none of their productions on the final day of training (two used the case marker in every production and three never used it). Thirty-three participants completed experiment 2, with four participants excluded for failing to achieve a 70% comprehension accuracy requirement and nine excluded for using case marking in all or none of their productions (seven used the case marker in every production, two never used it). Thus, productions from 20 participants were analyzed in each experiment.

Procedure. Participants visited the laboratory four times, each visit on a separate day with at most one day between the visits. During each visit, participants saw a mixture of exposure and test blocks. There were two types of exposure blocks and two types of test blocks:

Noun exposure. Participants viewed static pictures of people and objects one at a time and heard their names in the artificial language (30 trials total). The initial exposure was followed by a series of short vocabulary tests where participants were asked to choose the matching picture (out of two) for the character name they heard and to name the character shown on the screen. Feedback on performance was provided after each trial.

Sentence exposure. Participants viewed 80 short computer-generated videos depicting transitive actions (one at a time) and heard an accompanying sentence describing the event in the artificial language. Participants were instructed to repeat each sentence aloud to facilitate learning.

Comprehension test. In each of 80 trials, participants heard a novel sentence in the language, accompanied by two static pictures of the referents described in the sentence, and were asked to identify the doer of the action.

Production test. Participants were shown a novel transitive scene (80 trials total) and were instructed to describe it in the language learned during the experiment, using a provided verb prompt.

On day 1, participants completed the following blocks: noun exposure, sentence exposure, noun exposure, and a comprehension test. On days 2–4, the sequence of blocks was the same as on day 1 followed by a final production test block (see also Fig. 1).

Input Languages. The input languages of both experiments contained 8 verbs and 15 nouns. Both input languages had flexible constituent order: SOV order was dominant and occurred in 60% of the input sentences; OSV order was the minority constituent order and occurred in 40% of the input sentences. Both languages had optional case marking but differed in whether the grammatical object (experiment 1) or subject (experiment 2) was optionally case-marked. The case marker was always “kah,” and it always followed the noun whose case it marked. The frequency of case marking was identical across the two experiments: 60% of objects (experiment 1) or subjects (experiment 2) were overtly case-marked and 40% were not. By design, case marking was always independent of animacy (i.e., animate and inanimate nouns were equally likely to be case-marked). Case marking did vary by constituent order: 50% of OSV sentences were case-marked, and 67% of SOV sentences were case-marked. See *SI Text* for complete details.

In both experiments, the actions and the verbs were compatible with any of the referents being either the agent or undergoer. There were no differences in subcategorization frequencies between the verbs. That is, the frequency with which a noun was the subject or object did not differ between the verbs. The referents of the nouns and the actions referred to by the verbs differed, however, between the two languages (the former by design, the latter by necessity because the inanimate agents used in experiment 2 strongly constrained the choice of compatible actions).

ACKNOWLEDGMENTS. We thank C. Dolan and V. Choi for help with collecting subject data, C. Kurumada for suggesting optional case-marking as a test case for our hypothesis, and J. Elman and J. Trueswell for additions to the *Discussion* section. This work was supported, in part, by National Science Foundation Grants BCS-0845059 and IIS-1150028, an Alfred P. Sloan fellowship (to T.F.J.), and National Institutes of Health Grants DC00167 and HD037082 (to E.L.N.).

- Chomsky N (1965) *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA).
- Fodor JD (2001) Setting syntactic parameters. *The Handbook of Contemporary Syntactic Theory*, eds Baltin M, Collins C (Wiley-Blackwell, Malden, MA), pp 730–767.
- Christiansen MH, Chater N (2008) Language as shaped by the brain. *Behav Brain Sci* 31(5):489–508.
- Deacon TW (1997) *The Symbolic Species: The Co-evolution of Language and the Brain* (W. W. Norton & Co., New York).

- Newport EL (1981) Constraints on structure: Evidence from American Sign Language and language learning. *Aspects of the Development of Competence*, The Minnesota Symposia on Child Psychology, ed Collins WA (Erlbaum, Hillsdale, NJ), Vol 14, pp 93–124.
- Slobin DI (1973) Cognitive prerequisites for the development of grammar. *First Language Acquisition: The Essential Readings*, eds Lust B, Foley C (Wiley, New York).
- Bever TG (1970) *The Cognitive Basis for Linguistic Structures* (Wiley, New York), pp 279–362.

8. Hawkins JA (2007) Processing typology and why psychologists need to know about it. *New Ideas Psychol* 25(2):87–107.
9. Bates EA, MacWhinney B (1982) Functionalist approaches to grammar. *Language Acquisition: The State of the Art*, eds Wanner E, Gleitman L (Cambridge Univ Press, Cambridge, UK), pp 173–218.
10. Givón T (1991) Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Stud Lang* 15(2):335–370.
11. Zipf GK (1949) *Human Behavior and the Principle of Least Effort* (Addison-Wesley Press, Cambridge, MA).
12. Jaeger TF, Tily H (2011) On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdiscip Rev Cogn Sci* 2(3):323–335.
13. Dryer MS (1992) The Greenbergian word order correlations. *Language* 68(1):81–138.
14. Dryer MS (2007) Word order. *Clause Structure, Language Typology and Syntactic Description*, ed Shopen T (Cambridge Univ Press, Cambridge, UK), Vol 1, pp 61–131.
15. 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.
16. Croft W, Bhattacharya T, Kleinschmidt D, Smith DE, Jaeger TF (2011) Greenbergian universals, diachrony and statistical analyses. *Linguistic Typology* 15(2):433–453.
17. Tily H, Jaeger TF (2011) Complementing quantitative typology with behavioral approaches: Evidence for typological universals. *Linguistic Typology* 15(2):497–508.
18. Hudson Kam CL, Newport EL (2009) Getting it right by getting it wrong: when learners change languages. *Cognit Psychol* 59(1):30–66.
19. Hudson Kam CL, Newport EL (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang Learn Dev* 1(2): 151–195.
20. Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274(5294):1926–1928.
21. Christiansen MH (2000) Using artificial language learning to study language evolution: Exploring the emergence of word order universals. *The Evolution of Language: Third International Conference* (Ecole Nationale Supérieure des Telecommunications, Paris), pp 45–48.
22. Culbertson J, Smolensky P, Legendre G (2012) Learning biases predict a word order universal. *Cognition* 122(3):306–329.
23. Finley S, Badecker W (2008) Substantive biases for vowel harmony languages. *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed Bishop J (Cascadia Proceedings Project, Somerville, MA), pp 168–176.
24. Hupp JM, Sloutsky VM, Culicover PW (2009) Evidence for a domain-general mechanism underlying the suffixation preference in language. *Lang Cogn Process* 24(6):876–909.
25. MacWhinney B (1983) Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. *J Psycholinguist Res* 12(5):467–478.
26. Morgan JL, Meier RP, Newport EL (1987) Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognit Psychol* 19(4):498–550.
27. Newport EL, Aslin RN (2004) Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognit Psychol* 48(2):127–162.
28. Tily H, Frank MC, Jaeger TF (2011) The learnability of constructed languages reflects typological patterns. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (Cognitive Science Society, Austin, TX), pp 1364–1369.
29. 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):10681–10686.
30. Smith K, Wonnacott E (2010) Eliminating unpredictable variation through iterated learning. *Cognition* 116(3):444–449.
31. Shannon C (1948) A mathematical theory of communications. *Bell Syst Tech J* 27(4): 623–656.
32. Aylett MP, Turk A (2004) The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech* 47(Pt 1):31–56.
33. Genzel D, Charniak E (2002) Entropy rate constancy in text. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Stroudsburg, PA), pp 199–206.
34. Florian Jaeger T (2010) Redundancy and reduction: Speakers manage syntactic information density. *Cognit Psychol* 61(1):23–62.
35. Levy R, Jaeger TF (2007) Speakers Optimize Information Density Through Syntactic Reduction. *Advances in NIPS* (MIT Press, Cambridge, MA), pp 849–856.
36. Ferrer i Cancho R, Diaz-Guilera A (2007) The global minima of the communicative energy of natural communication systems. *J Stat Mech* 2007:P06009.
37. Piantadosi ST, Tily H, Gibson E (2011) Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA* 108(9):3526–3529.
38. Qian T, Jaeger TF (2012) Cue effectiveness in communicatively efficient discourse production. *Cogn Sci* 36(7):1312–1336.
39. van Son RJJH, van Santen JPH (2005) Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Commun* 47(1-2):100–123.
40. Piantadosi ST, Tily H, Gibson E (2012) The communicative function of ambiguity in language. *Cognition* 122(3):280–291.
41. Manin D (2006) Experiments on predictability of word in context and information rate in natural language. *J Inform Processes* 6(3):229–236.
42. Maurits L, Perfors A, Navarro D (2010) Why are some word orders more common than others? A uniform information density account. *Adv Neural Inf Process Syst* 23: 1585–1593.
43. Mohanan T (1994) *Argument Structure in Hindi* (Center for the Study of Language and Information, Stanford Univ, Stanford, CA).
44. Bossong G (1985) *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen [Empirical Study of Universals. Differential Object Marking in Modern Iranian Languages]*. (Narr, Tübingen, Germany). German.
45. Aissen J (2003) Differential object marking: Iconicity vs. economy. *Nat Lang Linguist Theory* 21(3):435–483.
46. Silverstein M (1976) *Hierarchy of Features and Ergativity. Grammatical Categories in Australian Languages*, ed Dixon RMW (Australian Institute of Aboriginal Studies, Canberra), pp 112–171.
47. Lee H (2006) Parallel optimization in case systems: Evidence from case ellipsis in Korean. *J East Asian Linguist* 15(1):69–96.
48. Wonnacott E, Newport EL, Tanenhaus MK (2008) Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognit Psychol* 56(3):165–209.
49. Yarbus AL (1967) *Eye Movements and Vision* (Plenum Press, New York).
50. Malchukov AL (2008) Animacy and asymmetries in differential case marking. *Lingua* 118(2):203–221.
51. Merlan F (1982) *Mangarayi* (North-Holland, Amsterdam).
52. Mosel U, Hovdhaugen E (1992) *Samoan Reference Grammar* (Scandinavian Univ Press, Oslo).
53. Arnold JE, Wasow T, Asudeh A, Alrenga P (2004) Avoiding attachment ambiguities: The role of constituent ordering. *J Mem Lang* 55(1):55–70.
54. Ferreira VS, Dell GS (2000) Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognit Psychol* 40(4):296–340.
55. Jaeger TF (2006) Redundancy and syntactic reduction in spontaneous speech. PhD dissertation (Stanford Univ, Stanford, CA).
56. Wasow T, Arnold J (2003) Post-verbal constituent ordering in English. *Determinants of Grammatical Variation in English*, eds Rohdenburg G, Mondorf B (de Gruyter, New York), pp 119–154.
57. Ferreira V Avoid ambiguity! (If you can)(2006) *Center for Research in Language Technical Reports* (Center for Research Language, Univ of California, San Diego, CA) Vol 18, pp 3–13.
58. van Son RJJH, Pols LCW (2003) How efficient is speech? *Proc Inst Phon Sci* 25:171–184.
59. Lindblom B (1990) On the communication process: Speaker-listener interaction and the development of speech. *Augment Altern Commun* 6(4):220–230.
60. Christianson K, Hollingworth A, Halliwell JF, Ferreira F (2001) Thematic roles assigned along the garden path linger. *Cognit Psychol* 42(4):368–407.
61. Frank A, Jaeger TF (2008) Speaking rationally: Uniform information density as an optimal strategy for language production. *The 13th Annual Meeting of the Cognitive Science Society* (Cognitive Science Society, Austin, TX), pp 933–938.