

## **Biased statistical learning of closed-class items**

Heidi R. Getz & Elissa L. Newport

*Department of Neurology, Georgetown University Medical Center, Washington, DC, USA*

Corresponding author: Heidi Getz, [hrg2@georgetown.edu](mailto:hrg2@georgetown.edu), Department of Neurology,  
Georgetown University Medical Center, Washington, DC, USA

## **Biased statistical learning of closed-class items**

In natural languages, closed-class items predict open-class items but not the other way around. For example, in English, if there is a determiner there will be a noun, but nouns can occur with or without determiners. Here we asked whether statistical learning of closed-class items is also asymmetrical. In three experiments we exposed adults to a miniature language with the one-way dependency “if X then Y”: if X was present, Y was also present, but Y could occur without X. We created different versions of the language in order to ask whether learning depended on which category (X or Y) was an open or closed class. In one condition, X had the main properties of a closed class and Y had the main properties of an open class; in a contrasting condition, X had properties of an open class and Y had properties of a closed class. Learners’ exposure in these two conditions was otherwise identical. Learning was significantly better with closed-class X. Additional experiments demonstrated that it is the perceptual distinctiveness of closed-class items that drives learners to analyze them differently, and that the mathematical relationship between closed- and open-class items influences learning more strongly than their linear order. These results suggest that statistical learning is biased: learners privilege computations in which closed-class items are predictive of, rather than predicted by, open-class items. We suggest that the distributional asymmetries of closed-class items in natural languages—and perhaps the asymmetrical structure of linguistic representations—may arise in part from this learning bias.

Keywords: statistical learning; function words; learning biases; syntax; morphology

Word count: 11,578 (including abstract and references but not figure captions or appendix)

## Introduction

In natural languages an important contrast is between open-class lexical items—for example, nouns or verbs—and closed-class or function items—for example, *is* or *the*.<sup>1</sup> Open-class categories like noun or verb contain many members and typically carry the important lexical content of the sentence. In contrast, closed-class items, which are used to mark the grammatical functions of other words, are typically very short, few in number, are each used with high frequency, and occur in predictable positions in their phrases. For example, English marks definiteness with the article *the*, which is one of the most frequent words in the language and always occurs before its noun. There is wide variation in the distribution of functional items across languages: in contrast to English, definiteness in Amharic is marked as a suffix on lexical items in a particular structural position and can attach to nouns, adjectives, numerals, or even verbs depending on sentence structure (Kramer, 2010). The distribution of closed-class items is always predictable in certain ways, but learners must do a substantial amount of distributional analysis in order to learn the particular patterning of closed-class items in their language. The goal of the present paper is to explore the computational mechanisms that enable language learners to do this.

From previous research we know that closed-class items draw special attention from language learners. Infants can identify them on the basis of their distinctive phonological, prosodic and distributional properties such as their short duration, light syllable structure, and high frequency (Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999), and children begin to represent these items well before producing them (Shafer, Shucard, Shucard, & Gerken,

---

<sup>1</sup> The terms ‘functional item’ and ‘closed class’ are often used interchangeably. We adopt the terminology of closed and open classes because these terms more readily apply to our miniature languages.

1998; Shi, Werker, & Cutler, 2006). Early attention to closed-class items could facilitate other aspects of language acquisition. Research with adults and older children showed that artificial languages were more easily learned when they contained words that were relatively high frequency or in other ways like the closed-class items of natural languages (Braine, 1963; Morgan, Meier & Newport, 1987; Valian & Coulson, 1988). The authors of these studies argued that closed-class items focus learners' attention on grammatically important parts of the sentence—such as phrase boundaries, where these items tend to occur—facilitating the acquisition of grammatical structure (Valian and Coulson call this the Anchoring Hypothesis). Additional empirical support for this hypothesis has come from more recent research showing that infants use familiar closed-class items to segment new words, assign words to grammatical categories, and form early representations of word order (Shi, Marquis, & Gauthier, 2006; Shi & Lepage, 2008; Shi & Melançon, 2010; Höhle et al., 2004; Zhang, Shi, & Li, 2015; Gervain et al., 2008; Bernard & Gervain, 2012; de Carvalho, He, Lidz, & Christophe, 2019; Mintz, 2006; Marino, Bernard, & Gervain, 2020). Like adults, infants also use novel words that are high frequency or phonologically like closed-class items to organize words in an artificial language (Bernard & Gervain, 2012; Gervain et al., 2008; Marquis & Shi, 2012), indicating that infants do not need extensive experience with closed-class items to use them for learning.

All of these studies support a general version of the Anchoring Hypothesis: learners notice closed-class items early on because of their distinctive perceptual and distributional properties, and then use these items to guide subsequent distributional analysis. However, it is not clear what computational mechanisms underlie this learning process. What do learners actually do with closed-class items, once they have noticed them? Only a few statistical learning studies have hypothesized specific statistical computations that learners might be using to

acquire linguistic information (though none of these have focused on closed-class items). These studies have suggested, for example, that learners can compute transitional probabilities to find word boundaries (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and to acquire grammatical phrases (Saffran, 2001; Thompson & Newport, 2007). More recent work has examined the computations underlying the acquisition of grammatical categories (Reeder, Newport & Aslin, 2013; Schuler, Reeder, Newport, & Aslin, 2017) and productive rules (Schuler, Yang, & Newport, 2016). Despite this progress, we are only beginning to identify the computational mechanisms underlying many aspects of language acquisition. It thus remains a mystery how learners manage to sort out patterns as complicated as (for example) Amharic definiteness. What kind of computations would a learner need to perform in order to acquire this type of pattern?

Consider the statistical information about closed-class items that is present in learners' input. As already noted, closed-class items generally do not independently contribute much semantic meaning; rather, they specify the grammatical properties or grammatical function of the meaning-bearing elements (the lexical categories). This role gives closed-class items a highly predictable syntactic context. For example, *the* indicates that its noun refers to a specific referent identifiable in context and therefore must appear with a corresponding noun, never alone. In statistical terms, the probability of seeing a noun, given that there is a determiner, is 100%. The reverse is not true, however, since nouns can occur in a variety of grammatical contexts, with or without determiners.<sup>2</sup>

---

<sup>2</sup> In some cases, predictiveness goes both ways (e.g., in French, all non-proper nouns require determiners). Nonetheless, computing how often determiners are accompanied by nouns will *always* reveal a pattern, whereas the reverse computation only sometimes will. Thus analyzing closed-class items as predictive of open-class items is the most effective way to discover linguistic patterns.

The statistical asymmetry in the distribution of closed- and open-class items is especially interesting in light of the recent emphasis in syntactic theory on the role of functional categories in sentence structure (see Rizzi & Cinque, 2016, for discussion and historical context). Increasingly, linguists have argued that properties of closed-class items determine the behavior of other words in the sentence. This extends beyond the presence or absence of certain open-class categories to their positions in the sentence as well. To illustrate, consider the pattern of verb placement in French. Lexical verbs such as “eat” (*mang-*) can either precede or follow the negative marker (*pas*), depending on whether the verb is morphologically finite, as in *tu manges pas?* (“You’re not eating?”), or non-finite, as in *tu vas pas manger?* (“You’re not going to eat?”). Linguists capture this contingency between finiteness and verb position by positing that the abstract features of finite and non-finite morphemes are represented in different positions in the sentence. If there is finite morphology, there will be a verb and that verb will occur in the “finite position” (pre-negation). In this way the presence and location of verbs is determined by the kind of morphology that occurs in the sentence.

Of course, linguists’ analyses are intended to be formal mathematical descriptions of sentence structure and not claims about the psychological representation of sentences. However, this kind of analysis demonstrates an important empirical point: *regularities of word order and word form can be stated as restrictions on the distribution of closed-class items*. Consider now the problem of distributional learning. One way to begin learning, given this view from syntactic theory, would be to identify closed-class items (for example, based on their salient perceptual properties) and then learn the distribution of open-class items relative to them. Because this distribution is asymmetrical—closed-class categories always predict but are not predicted by open-class categories—the most efficient computations that learners perform would also be

asymmetrical. Learners need to learn what a closed-class item *predicts*—the presence of other categories, the placement of words, and so on—but they need not expend any effort finding distributional patterns that a closed-class item is *predicted by*, because there are none.<sup>3</sup>

Here we explore the possibility that learners favor statistical learning computations in which closed-class items are *predictive of* open-class items over computations in which they are *predicted by* open-class items. In Experiment 1 we exposed adults to a miniature language containing a one-way grammatical dependency between two form-class categories, X and Y. When an X word was present, a Y word always had to be present as well, but Y words could occur with or without X words (“if X then Y”). This is mathematically like the relationship between determiners and nouns in English. In two contrasting conditions, we assigned different types of words to the X and Y categories. In one condition the predictive category (X) was a closed class (short, monosyllabic, and containing only one item *ka*), while in the other condition the predictive category was an open class (mono or disyllabic and containing three possible lexical items). Learning was better when X was closed-class, suggesting that statistical learning is biased: learners identify patterns where closed-class items are predictive of open-class items more readily than the reverse, indicating asymmetrical statistical computations. In Experiment 2 we asked whether this learning bias is driven by the distinctiveness of closed-class items. When

---

<sup>3</sup> Of course closed-class items do not appear randomly in sentences. Their presence is determined by the semantic meaning that the speaker wishes to express. The learner does eventually need to learn which meanings go with which forms, but this is a separate and somewhat uncorrelated problem. As the comparison between Amharic and English definiteness marking illustrates, learning that a given form means “definite” does not tell the learner where, distributionally, that form occurs, nor does learning the distribution of a form reveal its meaning (e.g., both definite and indefinite articles precede nouns in English). Both learning problems are important, but we are concerned here only with the distributional one.

the closed-class item was still high frequency but no longer phonologically distinctive, the group differences diminished but did not entirely disappear, consistent with the hypothesis that learners analyze closed-class items differently because they are perceptually distinctive. Finally, in Experiment 3 we demonstrate that the mathematical relationship between closed- and open-class items influences learning more strongly than their linear order. Together the results suggest that learners analyze dependencies in certain biased ways, more readily learning the kinds of patterns that exist in natural languages.

In the Discussion we ask why learners are biased in this way. We do not mean to suggest that they know innately about languages in particular; rather we suggest that their computational biases—which may arise from cognitive constraints outside of language—shape languages and cause them to be structured in this way.

Before proceeding, it is important to clarify an aspect of our experimental design. In contrast to many other studies from our lab, the artificial language that we created for these experiments is not very language-like. The experiments are focused on a specific computational question about how learners analyze closed and open-class items. To test our hypothesis, it was necessary to design a language that could only be learned by computing the precise mathematical relationship between two specific terms ( $X$  and  $Y$ ). For this reason,  $X$  was the only category whose distribution was constrained with respect to other words; all other words in the language appeared and disappeared freely, which is unlike the more extensively constrained sentence structures of natural languages. This experimental design allowed us to test empirically whether learners' computational analyses are biased in a certain way. If the results of these experiments do reveal such a bias, there is motivation to explore how this bias affects the acquisition of more naturally structured languages—a line of work that is in progress.



## **Experiment 1**

As described above, in our first experiment we presented adult participants with a miniature language in which class X predicted class Y but not the reverse. In two conditions we varied the characteristics of class X in order to see whether learning was better when class X items were more like a closed class in natural languages.

### ***Method***

#### *Participants*

Sixteen adults from the Georgetown University community (age 18-27, mean = 20.4) were randomly assigned to one of two conditions (8 participants per condition; see below). Data for one additional participant did not save due to an error. Each participant received \$10 for participating in this study.

#### *Description of the miniature language*

The design of the language is summarized in Figure 1. The word order was AXYBC, where each letter represents a form-class category. All categories were optional, with the constraint that up to three categories could be omitted per sentence (i.e. every sentence must have at least two words). The fixed and consistent rule of the language was that if X was present, Y had to be present (“if X then Y”). Thus every sentence with X also contained Y, but sentences with Y did not have to contain X. This relationship is analogous to the one between determiners and nouns in English. Note that this dependency is defined in terms of the conditional relationship, not the linear order, of X and Y. In Experiment 1 X preceded Y, but in Experiment 3 X followed Y; this difference in linear order does not change the conditional relationship between the two terms. The grammar in Experiment 1 generates 19 different sentence structures, ranging from two to five words long.

< Figure 1 here >

We created different versions of the language in order to ask whether learning the conditional relationship between X and Y depended on which of these terms was a closed-class or an open-class category. None of the words had any meaning; this contrast was defined by the number of words in each category and the phonological properties of those words.

In one condition, X was closed class and Y was open class (Closed X). The closed-class category contained a single item *ka*, which had several properties common to closed-class items in English: it was short, lacked a coda or consonant clusters, and was high frequency by virtue of being the only member of its form class. In contrast, each open-class category contained three words that were either mono- or disyllabic forms. All words in the language, including the closed-class item, carried stress (i.e., *ka* was not prosodically dependent on any other item). In the Closed X condition, the X category contained *ka* and Y contained *lapal*, *tombur*, and *zup*. The closed-class item *ka* predicted any of these three open-class items.

In a contrasting condition, X was open class and Y was closed class (Open X). The items in the X and Y categories were the exact opposite of the Closed X condition: X contained *lapal*, *tombur*, and *zup* and Y contained *ka*. In this condition the closed-class item *ka* is predicted by each of three open-class items (rather than being the predictor of them).

Other than the specific lexical items in the X and Y categories, the two versions of the language were identical. In both versions, sentences with X must also contain Y, while sentences with Y may or may not contain X. Because either X or Y is *ka*, learners in both conditions had an “anchor” for their distributional analyses. In both conditions, the predictive category (X) comes before the category it predicts (Y); this linear order was like participants’ native language, English, where (for example) determiners precede nouns. (In Experiment 3 we reversed the

linear order such that the predictive category came last, as in languages like Japanese.) At a lexical level, in both conditions the dependency involved exactly one closed-class item and three open-class items; acquiring the dependency required computing exactly three word-level forward transitional probabilities (either X1-Y1, X1-Y2, X1-Y3 in the Closed X condition or X1-Y1, X2-Y1, X3-Y1 in the Open X condition).

Because these languages are identical except for the closed/open class contrast for X and Y, learning outcomes will differ only if learners' computational analyses treat closed-class and open-class items differently. If learners preferentially analyze closed-class items as predictive, they should more easily learn "if X then Y" in the Closed X condition, where *ka* predicts an open-class category; but they should struggle in the Open X condition, where *ka* is predicted by an open-class category. Alternatively, if learners analyze closed-class and open-class items similarly, learning outcomes will be equivalent across conditions.

### *Materials*

Words were selected from a lab corpus based on their length and syllable structure in order to create a distinction between closed and open-class items as described above. Individual words were recorded in list intonation by a female native speaker of English and then concatenated to form sentences with 50 msec of silence between words.

We generated a 38-sentence exposure set by selecting two sentence types for each of the 19 possible structures of the language. The sentence structures were the same for all of the experiments in this paper, but the actual sentence strings differed across conditions and experiments according to the lexical items that were in the X and Y categories (sentences are included in the Appendix). The 38-sentence exposure set was presented 16 times in a 1-back task (see Procedure). Sentences were presented in a different order on each loop.

### *Procedure*

Participants learned the language through a computer game programmed in PsychoPy (Peirce, Gray, Simpson, et al., 2019). Instructions were spoken aloud by a robot named “Bot,” who told participants to listen as an alien named Zooma practiced saying sentences in a made-up language spoken by aliens on a planet to which she was traveling. Instructions were synthesized in MacInTalk using the Alex voice. Then Zooma began flying around the screen saying sentences in the language. After each sentence, participants pressed a button to indicate whether Zooma had just repeated herself. This 1-back task was designed to ensure that participants paid attention to the sentences but did not otherwise provide feedback or explicit information about the structure of the language. As the experiment progressed, the background image changed so that Zooma appeared to be traveling from Earth through various galaxies to the new planet. When she arrived on the new planet, the test began. Bot explained that Zooma would try to say each sentence two different ways, and the participants’ job was to decide which one was better. They pressed a button on the screen to indicate their choice. The entire experiment lasted approximately 45 minutes.

### *Test*

Learning was measured as accuracy on a two-alternative forced-choice (2AFC) test. The target choice on each trial was always a complete grammatical sentence. The alternative was identical to the target except that one word was changed, or the words were the same but in a different order. Test items were of several types.

**XY test items, controlled** (4 items). The primary goal of the test was to ask whether participants had acquired a very specific piece of knowledge: the conditional relationship between X and Y. In order to answer this question it was important to create test items on which

all other distributional properties (e.g., bigram frequency) were controlled. Only two types of test items could be carefully controlled in this way, so items with confounds were also included but scored separately (see below). The first type of well-controlled XY test item served as a constituency test, with participants choosing between XY and YB (“Constituency Test”). Both choices are legal two-word sequences and complete sentences, have the same relative frequency in learners’ input (each bigram occurred in 16 out of 38 exposure sentences, including twice each as a complete sentence), and are medial bigrams in the basic sentence structure (AXYBC). However, X and Y are related grammatically whereas Y and B are not. A preference for XY would indicate that participants represent this grammatical relationship. A second type of XY test item tested whether participants learned that X predicts Y, but not the reverse (“Dependency Test”). The alternatives were AY and AX. Again, both choices are legal two-word sequences and occurred in learners’ input with the same relative frequency (each bigram occurred in eight of 38 exposure sentences). In addition, the two sequences had exactly the same forward transitional probability (.36). However, only AY is a complete sentence. AX is a grammatical sequence but not a complete sentence, since it contains X but not Y. If participants have learned that X predicts Y (but not the reverse), they should prefer AY over AX. Note that while participants in the Closed X condition could succeed on the first item type by selecting the choice that contains *ka*, the same strategy would not work on the second item type, since there the choice with *ka* is incorrect.

**XY test items, with confounds** (4 items). As described above, it was critical to ensure that the target and foil on XY test items were matched on various distributional properties. However, because only a small number of test items could meet this criterion, we also included some items where distributional properties could not be matched. These had the same structure

as the constituency test described above: participants chose between XY and BC or XY and AX. On these trials, the target and foil differed in their distribution in various ways (e.g., XY is more frequent than either BC and AX, but AX and BC are both edge bigrams in the basic sentence structure while XY is medial). Therefore participants might or might not prefer the target choice on these for a variety of reasons unrelated to the main hypothesis of this study. Nonetheless, results for these items can contribute to a general understanding of how well participants learned each language overall.

**Basic order** (4 items). On items testing basic word order, participants chose between AB or AC and an ungrammatical alternative containing the same words in reverse order (that is, \*BA or \*CA). These items were included as a basic measure of learning; we expected both conditions to succeed on these items.

**Fillers** (6 items). We included filler trials to balance the frequency with which targets and foils for the XY Dependency trials appeared on the test. This ensured that participants could not succeed on the critical test trials by selecting the response they had heard most frequently on the test overall (which without these filler items would have been XY). These items were not scored.

**Longer sentences** (20 items). With longer sentences, it is even more difficult to isolate structural properties of the language for test. For example, if participants are asked to choose between AYB and \*AXB, a preference for AYB could reflect either knowledge of the XY dependency or a preference for the bigram YB over \*XB. This was problematic. If learners are computationally biased in the way we have suggested, learners in the Open X condition are predicted to focus on their closed-class item (Y) and the categories with which it most reliably occurs (B and C, each with a conditional probability of .42; the forward transition probabilities of YB and YC are .53 and .27, respectively). In other words, learners might perform well on this

test for different reasons: learners in Closed X might reject \*AXB based on knowledge of the XY dependency, while learners in Open X might prefer AYB based on recognition of the YB bigram. Thus these trials also cannot be used to determine whether participants acquired the precise mathematical relationship between X and Y. However, these items do add information about how well each set of participants acquired the language in general. Items testing longer sentences were always presented at the end of the test, after all of the item types described above.

### *Analysis*

Accuracy on the 2AFC test was measured as the proportion of trials on which participants chose the target sequence. Note that in some cases (e.g., the XY constituency test) both choices are complete sentences. If participants show a preference, this indicates that one choice is more strongly represented in their knowledge of the language.

Results were analyzed statistically using logistic regression models run in R (Version 3.6.1, R Core team, 2019). For each model described in the paper, we first generated a mixed-effects logistic regression model using the lme4 package (Bates et al., 2015) with condition (Closed X (reference level) vs. Open X) as a fixed effect and random intercepts for participants and items. In most cases, including random by-participant and by-item intercepts led to convergence errors, so in these cases the results were re-analyzed using standard fixed-effects logistic regression models.

## ***Results***

### *XY test items*

The results for the critical test items in Experiment 1 are presented in **Figure 2**. Performance on these items was analyzed in a standard logistic regression model with condition (Closed X vs. Open X) as a fixed effect. (Mixed-effect models with random by-participant and/or by-item

intercepts did not converge.) The proportion of trials on which participants chose the target sequence was significantly higher for Closed X than Open X (Closed X:  $M = .84$ ,  $SD = .13$ , Open X:  $M = .53$ ,  $SD = .25$ ,  $\beta = -1.56$ ,  $SE = 0.60$ ,  $z = -2.59$ ,  $p < .01$ ). The strong performance of Closed X was also robust at the individual level: all eight participants in this condition reached 75% accuracy or higher. **Table 1** presents the results separately for the two types of test items; performance was similar across item types.

The structure of the XY dependency was identical across the two conditions, since class X always perfectly predicts class Y. Yet learning outcomes differed: the “if X then Y” rule was more easily learned when X was a closed class than when X was an open class and Y was a closed class. These results support our hypothesis that learners analyze closed-class items as predictive of open-class items more readily than the reverse.

### *Basic Structure*

As a general measure of learning, we also tested knowledge of basic sentence structure (**Table 2**). On these items, participants in both conditions chose the target sequence more often than the foil (Closed X:  $M = .94$ , Open X:  $M = .84$ ). These results indicate that both groups of participants acquired basic knowledge of the language.

### *Additional items*

Results for the remaining test trials are presented in Table 2. The distributional confounds in these remaining items (see Methods) mean that participants could perform well on these items without knowing the precise mathematical relationship between X and Y. Therefore, results on these items do not provide evidence for or against our hypothesis. On these items, participants in both groups performed well, choosing XY more often than AX/BC (Closed X:  $M = .72$ , Open X:  $M = .66$ ) and preferring longer grammatical sentences to longer ungrammatical sentences



(Closed X:  $M = .66$ , Open X:  $M = .65$ ). These results indicate that participants in both groups acquired some knowledge of the language.

< **Figure 2 here** >

< **Table 2 here** >

### ***Discussion***

In Experiment 1, adults were exposed to a miniature language with a single fixed and consistent rule: if X was present, Y was also present (“if X then Y”). We created two versions of the language, varying whether X or Y was a closed class. When X was a closed class and Y was open, participants easily acquired the XY rule. However, when X was open and Y was closed, learning was significantly worse. These results suggest that, during statistical learning, learners favor computations in which closed-class items are predictive—the same computations that are most relevant for acquiring natural language dependencies.

In the Introduction we hypothesized that learners might analyze closed-class items in this way because they are distinctive. That is, learners notice closed-class items first, hence these items come to serve as the constant terms around which to organize knowledge of other patterns. To test this aspect of the hypothesis, we did a second experiment in which the closed-class item was less distinctive (but still high frequency). We predicted that learning in the Closed X condition of this experiment would be reduced compared to Experiment 1.

### **Experiment 2**

## ***Method***

### *Participants*

Sixteen adults from the Georgetown University community (age 18-22, mean = 19.5) were randomly assigned to the two conditions (8 participants per condition). Each received \$10 for participating in this study.

### *Description of the miniature language*

The language in Experiment 2 was identical to Experiment 1 except that the closed-class item was less distinctive. Here the closed-class item (*daygin*) was disyllabic, carried initial stress, and had a closed final syllable, making it phonologically like the open-class words in the language; its only distinguishing property was its high frequency. If distinctiveness of *ka* drove learning outcomes in Experiment 1, learning should be weakened in Experiment 2. However, since *daygin* is still high frequency, learning may still be better for Closed X relative to Open X.

## ***Results***

Part of our hypothesis was that learners analyze closed-class items differently because they are distinctive. In Experiment 2 we tested this by making the closed-class item less distinctive: there was one closed-class item that was high frequency but phonologically like the open-class words in the language. The main results are shown in Figure 2. Results are broken down by test item type in Table 1 and Table 2.

Performance on the critical test items was analyzed with a standard logistic regression model with condition (Closed X vs. Open X) as a fixed effect. In contrast to Experiment 1, where there was a significant effect of condition, in Experiment 2 the effect of condition was marginal (Closed X:  $M = .66$ ,  $SD = .23$ , Open X:  $M = .44$ ,  $SD = .22$ ,  $\beta = -.90$ ,  $SE = .52$ ,  $z = -1.74$ ,  $p = .08$ ). The intercept of this model was also marginal, indicating that performance for the two groups

combined was not significantly above chance ( $\beta = .65$ ,  $SE = .37$ ,  $z = 1.74$ ,  $p = .08$ ). To directly compare learning in Experiments 1 and 2, we ran an additional model specifying fixed effects for condition, Experiment, and their interaction. There was a main effect of condition ( $\beta = -1.56$ ,  $SE = .60$ ,  $z = -2.59$ ,  $p < .01$ ), a marginal effect of Experiment ( $\beta = -1.04$ ,  $SE = .61$ ,  $z = -1.70$ ,  $p = .09$ ), and no interaction ( $\beta = .66$ ,  $SE = .79$ ,  $z = .84$ ,  $p = .40$ ).

This pattern of results is consistent with our hypothesis. Learning was somewhat weaker in Experiment 2, as reflected in numerically lower performance and marginal effects in the logistic regression model for that experiment. This suggests that making the closed-class item less distinctive—i.e., changing it from *ka* to *daygin*—weakens learning, as we predicted. Interestingly, however, this manipulation did not entirely eliminate the effect of condition. (If it had, we would have observed a Condition x Experiment interaction.) This suggests that the presence of a single high-frequency item—even if the item has no other distinguishing properties—begins to introduce some asymmetry in learners' distributional analyses, somewhat facilitating discovery of patterns in which the high-frequency item predicts other elements.

### ***Discussion***

Experiment 2 demonstrated that the distinctiveness of closed-class items drives learners to analyze them in a certain way. When we changed the closed-class item from *ka* to *daygin*, learning in the Closed X condition was only marginally better than in the Open X condition, whereas this difference in Experiment 1 was much greater. This is exactly what we would expect if the learning asymmetry we observed in Experiment 1 was driven by the distinctiveness of closed-class items. In Experiment 2 the closed-class item was still distinctive (due to its high frequency), but not strongly so, with the consequence that learning was still better for Closed X relative to Open X, but only a little bit. Thus the results of both Experiment 1 and Experiment 2

are consistent with our hypothesis that learners analyze closed-class items in certain biased ways and that this bias is driven by the distinctive phonological and distributional properties of those items.

*A simpler explanation: Linear order effects*

In Experiments 1 and 2, in the Closed X version of the language, the closed-class item came before the open-class item in the XY sequence ( $ka_X \text{ tombury}$ ; Figure 1). This matches the order of function words and their complements in our participants' native language, English (cf. “the dog”), and contrasts with the word-order patterns of head-final languages such as Japanese. Therefore it is possible that the effects we have seen so far—better learning for Closed X in Experiments 1 and 2—might not be due to our hypothesized account but could simply arise from native speakers of English being better at acquiring dependencies where the closed-class item comes first, as in English. Of course, English also has suffixes—bound morphemes that occur after their stems, as the plural marker does in “dogs”—and at least one artificial language study has found that English speakers are better at learning the category membership of words when categories are cued with suffixes relative to prefixes (St Clair, Monaghan, & Ramscar, 2009). The affixes in that study were similar in form to our closed-class item *ka* (all were high-frequency CV sequences). Based on that study, one could argue that the word order of the language in Experiments 1 and 2 actually favored the Open X condition, not the Closed X condition.

However, other evidence suggests that there is still reason to be concerned about the role of English transfer in explaining our results. Studies by Gervain and colleagues (Gervain et al. 2008, Bernard & Gervain 2012, Gervain et al. 2013) suggest that the word order of one's native language shapes adults' and infants' expectations about how words in a miniature language are

grouped. For example, Italian infants expect word pairs to have a frequent word first, while Japanese infants expect word pairs to have a frequent word last (Gervain et al., 2008). This expectation alone cannot explain our results. If participants were simply selecting test sequences in which frequent items come first, participants in the Open X condition would have preferred YB (“ka<sub>Y</sub> flugit<sub>B</sub>”) to XY (“lapal<sub>X</sub> ka<sub>Y</sub>”). This is not what we found: the Open X condition was at chance on these items in both Experiment 1 and Experiment 2.<sup>4</sup> However, participants’ expectations about word groupings may have influenced the results in other ways. In particular, learners may use the grouping structure of a language to identify the domains within which grammatical dependencies may hold (Morgan et al., 1987). The expectation that high-frequency elements mark the beginnings of phrases may have led participants in Closed X to search for patterns within sequences beginning with X, which always included Y (e.g. XY, XYB...), whereas this expectation may have led participants in Open X to search for patterns within sequences beginning with Y, which never included X (e.g. YB, YBC...) This could have produced better learning of the XY dependency for Closed X without a computational bias to analyze closed-class items in a particular way.

Experiment 3 was designed to assess this alternative explanation. We changed the word order of the language so that Y preceded X. Now the Closed X condition is superficially more like Japanese (frequent element last: tombur<sub>Y</sub> ka<sub>X</sub>), whereas the Open X condition is

---

<sup>4</sup> Interestingly, we did see a slight preference for YB in the Open X condition in pilot studies run on Mechanical Turk. We also saw a corresponding preference in the Open X condition of Experiment 3. There are multiple ways to account for this pattern of results, including through our original hypothesis; see the Discussion of Experiment 3. But in any case, even if some participants are using a linear-order strategy on the test, there might still be a computational bias that affects learning independently of linear order. Experiment 3 was designed to investigate this possibility.

superficially like English (frequent element first:  $kay\ tombur_X$ ). English-based expectations about word groupings would highlight the sequences containing both X and Y in the Open X condition, theoretically providing an advantage for learning the XY dependency. In contrast, in the Closed X condition these expectations would highlight sequences containing X but excluding Y, theoretically putting these learners at a disadvantage for learning the XY dependency. If the higher performance for Closed X in Experiments 1 and 2 reflects transferred expectations from English, then the results of Experiment 3 should be opposite to those of Experiment 1, with better learning for Open X. However, if learning outcomes depend on the structural relationship between closed-class and open-class items, results should be qualitatively similar to the results of Experiment 1: Closed X should learn “if X then Y” and Open X should fail. This is the outcome we predict.

Note that our hypothesis does not require learning in the Closed X condition of Experiment 3 to reach the same absolute level as in the Closed X condition of Experiment 1. If learning of Closed X is worse in Experiment 3 than in Experiment 1, this would simply indicate that English-speaking adults are better at learning a within-phrase dependency in a miniature language when the phrase begins with a frequent word, as in English. This would not be surprising in light of the evidence reviewed above. The question in Experiment 3 is whether transfer from English can *fully account* for the better performance of Closed X in Experiments 1 and 2. The important prediction is that the relative ease of learning Closed X over Open X should be the same as in previous experiments, even with non-English word order.

### **Experiment 3**

## *Method*

### *Participants*

Sixteen adults from the Georgetown University community (age 18-28, mean = 21.1) were randomly assigned to one of the two conditions (8 participants per condition). One additional participant did not complete the experiment due to a technical error. Participants received \$10 for participating in this study.

### *Description of the miniature language*

We used the same miniature language as Experiment 1, except that the word order was AYXBC instead of AXYBC. Here Y came before X in the linear order of the language, but X still predicted Y rather than the reverse ( $Y \leftarrow X$ ). This is like the relationship between determiners and nouns in a head-final language like Japanese. As in Experiment 1, we included two conditions that maintained the linear order of Y and X but differed in which member of the dependency (X or Y) was closed-class (**Figure 1**). The closed-class item was always *ka*.

### *Test*

The 2AFC test was modified slightly to take into account the linear order change. On the “if X then Y” dependency test, participants were now asked to choose between YX and XB instead of XY and YB. These are the two medial bigrams in the new basic sentence structure and, like XY and YB in Experiment 1, overlap by one word. These two sequences have nearly the same forward transitional probabilities (YX: .53, XB: .50) but different backwards transition probabilities (YX: 1.0, XB: .36) and different frequencies (YX: 16, XB: 8). The other trial type on the “if X then Y” dependency test (AY vs. \*AX) stayed the same. However, whereas both of these sequences were legal in Experiment 1 (and had the same forward transition probabilities), \*AX is not a legal sequence in Experiment 3, since X cannot be preceded by any item other than

Y. Based on the distributional properties of these test items, it should be relatively easier in Experiment 3 than in Experiments 1 and 2 for both conditions to select the target item on the test.

### **Results**

This experiment was designed to ask whether the learning advantage for Closed X in previous experiments is attributable to the fact that the closed-class item (*ka* or *daygin*) occurred first in the XY sequence. We tested learning of the same dependency, “if X then Y”, when X occurred after Y in the linear order of the language. The Closed X condition now has *ka* occurring at the end of the YX sequence, making that condition superficially unlike English and more like Japanese.

Results are presented in **Figure 1**. To understand the effect of linear order on learning, we ran two logistic regression models. First, we analyzed just the results of Experiment 3 in a mixed model with condition (Closed X vs. Open X) as a fixed effect, and random intercepts for individual test items. (The model did not converge with random by-participant intercepts.) The effect of condition was not statistically significant (Closed X:  $M = .72$ ,  $SD = .16$ , Open X:  $M = .56$ ,  $SD = .26$ ,  $\beta = -.70$ ,  $SE = .54$ ,  $z = -1.31$ ,  $p = .19$ ). In the second model we directly compared the results of Experiments 1 and 3, specifying fixed effects for condition (Closed X vs. Open X) and linear order (XY vs. YX) as well as an interaction between these terms. (A mixed model with random intercepts for participants and/or items did not converge.) The effect of condition was significant ( $\beta = -1.56$ ,  $z = -2.59$ ,  $p < .01$ ) but the effect of linear order was not ( $\beta = -0.75$ ,  $z = -1.20$ ,  $p = .23$ ). The interaction between condition and linear order was also not significant ( $\beta = 0.87$ ,  $z = 1.089$ ,  $p = .28$ ).

The results of the combined model for Experiments 1 and 3 suggest that the most important factor for learning is the mathematical relationship between closed-class and open-



class items—and not the linear order of these items. Consistent with this, the model analyzing just Experiment 3 found no learning advantage for Open X over Closed X (i.e. no significant effect of Condition). This refutes a possible explanation for the results of our previous experiments: that learning was better for Closed X only because that condition was superficially most like English. If this were true, the results of Experiment 3 would have been a mirror image of Experiment 1. Instead, the results of all three experiments are qualitatively quite similar (Figure 3), with higher accuracy for Closed X than Open X.

Quantitatively, the effect of condition was somewhat weaker in Experiment 3 relative to Experiment 1. Closed X dropped from 84% to 72% correct, while Open X improved slightly from 53% to 56%, and the effect of condition in the Experiment 3 model did not reach significance. This is not surprising and suggests that participants were at a slight disadvantage when learning a language with non-English-like word order (i.e., in the Open X condition in Experiment 1 and the Closed X condition in Experiment 3). However, the results of the combined model for Experiments 1 and 3 indicate that this effect of linear order is relatively small (it did not reach significance) and, most importantly, it is distinct from the much larger and significant effect of condition.

The results just reported are for two types of items: the constituency test (YX vs. \*XB) and the predictive direction test (AY vs. \*AX). In previous experiments, results were similar across item types (Table 1). The same was true for the Closed X condition of this experiment. However, the Open X condition of Experiment 3 showed a different pattern. These participants performed extremely well on the Constituency test (88% correct)—even outperforming the Closed X condition—but poorly on the Dependency test (31% correct). The results for the Dependency test are particularly striking because these items should have been easy: the foil

(\*AX) was not only an incomplete sentence, it was an illegal sequence (unlike in Experiment 1 where AX was an incomplete sentence but a legal sequence). There are several possible explanations for this pattern of results, one of which is provided by our original hypothesis. We consider these explanations in the following section.

### ***Discussion***

In Experiment 3, we changed the order of X and Y such that X came after Y (AYXBC). In the Closed X condition, the closed-class item *ka* came last in the critical sequence, matching the superficial position of closed-class items in head-final languages like Japanese and contrasting with participants' native language, English. Our question was whether the effect of condition would be reversed, with Open X now outperforming Closed X. This would indicate better learning for the language that was superficially most like English and would suggest that transfer from English explains the results of the previous experiments. This is not what we found. Instead, results were qualitatively similar to Experiments 1 and 2, with better learning for Closed X. In an analysis comparing the results of Experiments 1 and 3, only the class type of X—and not the linear order of X and Y—had a significant effect on test performance. These results suggest strongly that the mathematical relationship between closed-class and open-class items strongly influences learning—significantly more than the linear order of these items.

Although the overall results of Experiment 3 converge with prior experiments, the specific pattern of results across item types was slightly different. In previous experiments, performance was equivalent across item types (Table 1). In Experiment 3, however, learners in Open X performed very well on the Constituency test (preferring YX to \*XB) and poorly on the Dependency test (preferring \*AX to AY). It is possible that this pattern reflects random variation in participants' responses, especially given the small number of items of each type. It is also

possible that this pattern is a genuine reflection of participants' preferences. Our hypothesis, in its strongest form, actually predicts this pattern of results (and we had this in mind when designing the test). For this reason, we discuss this finding below. However, our general conclusions do not rest on how we interpret this small piece of the results.

There are two potential reasons that participants in the Open X condition of Experiment 3 might prefer YX to \*XB, but \*AX to AY. One is that they were using an English-based linear order heuristic on the test, choosing YX on the Constituency test because the frequent item (Y) came first, and \*AX on the Dependency test because the alternative had a frequent item last. We considered and rejected the idea that participants were using a linear-order heuristic when discussing the results of Experiments 1 and 2, because, in those experiments, there was no evidence that the Open X learners were using this strategy (see the Discussion of Experiment 2). However, there was a difference in the distribution of Y in Experiment 3 (due to the linear order of the language) that could have led the Open X participants in Experiment 3 to use this strategy, even if their counterparts in earlier experiments did not. In Experiment 3, Y occurred earlier in the sentence's linear order, making it much more likely to be the first word of the sentence. (In Experiments 1 and 2, 6/38 sentences began with Y; in Experiment 3, 14/38 sentences did.) This may have increased learners' attention to Y, and specifically to this element's frequent occurrence as the first item in a sequence, leading to an enhanced preference for sequences in which Y occurred first. If this explains the pattern of results for Open X, it would indicate that linear order is driving their preferences. Note that this explanation can only account for the specific pattern of performance across item types for Open X in Experiment 3. The general advantage of Closed X across experiments, including in Experiment 3, indicates that the

closed/open contrast has a significant effect on learning that is separate from the effect of linear order.

A second explanation for the pattern of results in the Open X condition in Experiment 3 comes from our original hypothesis. We have proposed that humans analyze dependencies in certain biased ways, more readily learning patterns in which closed-class items are predictive. Such a bias might lead learners in the Open X condition to initially analyze their closed-class item, Y, as predictive of some other item—such as X, the item that most frequently follows Y in this version of the language. Thus, a preference for YX (vs. \*XB) and \*AX (vs. AY) might reflect a generalization along the lines of “if Y then X”. This generalization is not consistent with learners’ input (the probability of X, given Y, is only .53), but it is consistent with the mathematical structure of dependencies in natural languages. Similar findings—where learners form a generalization that is not supported by the statistical structure of their language input—have been reported in other experiments where adult learners were exposed to miniature languages that were unnatural in certain ways (Hudson Kam & Newport, 2009; Fedzechkina, Jaeger, & Newport, 2012; Culbertson & Newport, 2017). There is parallel evidence from child learners, who appear even more likely than adults to generalize beyond their input when that input is inconsistent or comes from a miniature language that is unlike natural languages (Hudson Kam & Newport, 2005; Singleton & Newport, 2004; Culbertson & Newport, 2015). As with the linear-order heuristic, the reason we observed this pattern in Experiment 3 but not Experiment 1 could be that Y occurred at the beginnings of sentences much more frequently in Experiment 3. Elements that occur in first position are known to be particularly salient to learners (Ebbinghaus, 1885/1913) and this could have magnified the effects of a computational bias to analyze perceptually distinctive items as predictive.

It is worth noting that, when piloting Experiment 1 online through Mechanical Turk, we saw a pattern of results in the Open X condition that corresponded to the pattern observed in the Open X condition of Experiment 3 in this paper. That is, participants in Open X in the pilot version of Experiment 1 preferred \*YB to XY and \*AX to AY, resulting in significantly below-chance performance overall. (Participants in Closed X were reliably above chance in all of our pilot experiments, consistent with the experiments we report here.) As with the results for Experiment 3 in this paper, this pattern of performance could reflect either a linear-order strategy or an analysis in which the closed-class item Y is predictive of the item that comes after it most frequently (in this case B)—or both. Either of these biases may be more likely to manifest in situations where there is decreased learning of the actual statistics of the language, which we do sometimes observe when running studies online.

## **General Discussion**

In three experiments, we asked whether learners analyze closed-class and open-class items differently. We exposed participants to an artificial language with one consistent rule: when X was present, Y was also present (“if X then Y”). In Experiment 1 participants easily acquired the rule when X was a closed class but failed to acquire the same dependency when X was an open class and Y was a closed class. In that experiment, the closed class had several properties of closed classes in natural languages. When the closed class had no distinguishing properties other than high frequency (Experiment 2), group differences diminished but did not disappear, indicating that learning in Experiment 1 was partially driven by the distinctive perceptual properties of the closed class and partially driven by its frequency. In Experiment 3 we altered the linear order of the language so that Y came before X in the language’s linear order, but X’s presence still predicted Y’s presence in any given sentence. Results were qualitatively the same

as in the first two experiments, indicating that the primary determinant of learning is the mathematical relationship between closed and open classes not their linear order. Taken together, these results suggest that learners privilege computations in which closed classes predict open classes—the same computations that are most relevant for natural language dependencies.

It is important to note that, within each experiment, the Closed X and Open X conditions had exactly the same statistical evidence for the rule “if X then Y”. When class X was present, class Y was always also present in the same sentence (the conditional probability of Y, given X, was 1.0). Furthermore, the other statistical properties of the language were matched as well. Participants in contrasting conditions were exposed to the same number of lexical items, sentence structures, and sentence types. There were always exactly three item-level transitional probabilities from X to Y (or from Y to X), and these individual XY (or YX) bigrams occurred equally frequently across conditions (each occurred either 5 or 6 times in the exposure set of both conditions). Within each experiment, the linear direction of the XY dependency was the same, such that learners across the two conditions needed to compute transition probabilities in the same direction (forward for Experiments 1 and 2, backward for Experiment 3). Despite this mathematical equivalence, we observed a consistent learning advantage when X was a closed class. These results illuminate two important properties of learners’ distributional analyses.

First, learners’ distributional analyses are sophisticated. Our participants apparently were not simply storing high-frequency word sequences or chunks (cf. Ambridge, 2020; Freudenthal et al., 2007; McCauley & Christiansen, 2019; Perruchet, 2019). This approach would have produced equivalent learning across conditions, since the XY/YX bigrams were matched in frequency in the two versions of the language. The fact that we saw a consistent advantage for Closed X indicates that learners must have been tracking the conditional relationships among

word classes. The experiments in this paper thus add to a growing body of evidence that the computation of higher-order statistics such as conditional probability, and not simply storing high-frequency chunks, is important for acquiring language patterns.

Second, learners' distributional analyses are asymmetrical. Participants in our experiments apparently did not compute all of the pairwise conditional probabilities involving the closed-class item *ka* (or *daygin*). Like a chunk-based approach, this approach would have produced equivalent learning across conditions, since both versions of the language contained a consistent predictive dependency between a closed class and an open class. Instead, our results suggest that learners computed only a subset of the possible conditional probabilities involving the closed-class item: those in which the closed-class item was the constant (given) term. This finding sheds light on the statistical computations underlying anchored learning, as we discuss in the following section.

### ***Extending the Anchoring Hypothesis***

The original idea of the Anchoring Hypothesis (Valian & Coulson, 1988; Morgan, Meier, & Newport, 1987; see also Christophe, 1997 and Mintz, Newport, & Bever, 2002) was that, because closed-class items tend to occur at grammatically important points in the sentence such as phrase boundaries, focusing on them could help learners acquire grammatical structure. That is, closed-class items were thought to help primarily by indicating to learners the domains within which distributional analysis would be most productive. Our results add a computational component to this approach. Because closed-class items are noticed first, due to their distinctive phonological properties and their high frequency, these will be the constant terms in learners' computations; other patterns are learned and represented relative to them. Thus, in addition to highlighting important domains for distributional analysis, closed-class items may also help by

providing learners with a consistent term to use in their computations, structuring learners' analyses in a way that highlights natural language patterns.

### ***Effects on language representations***

A learning mechanism that operates in this way would ultimately represent a broad range of language patterns in terms of the distribution of a small set of closed-class items. As we pointed out in the Introduction, this is increasingly the way that language patterns are described by syntactic theory as well. A widely held view in contemporary linguistics is that this representational structure results from innate knowledge about how languages are organized. Our results are consistent with this claim: if learners know innately that sentences are structured around functional categories, this is the kind of computational analysis that they should do. Alternatively, however, there is another explanation that does not require such a strong claim about innate linguistic content. On this alternative account, learners do not know innately that there are functional categories and that the purpose of such categories is to organize sentence structure. Rather, learners attend to closed-class items because these items occur frequently and sound different from other elements. As a result, these items are learned early and are ultimately available to learners as units around which to organize their knowledge of sentences. If this is right, the privileged status of closed-class items in humans' mental representations is a consequence of the particular way in which learners analyze and learn language input.

### ***The role of anchored learning in natural language acquisition***

If closed-class items are so critical to learners' early distributional analysis, why do children omit these items from their early sentence productions? Some researchers have interpreted these omissions as evidence that children initially represent only open-class items (Brown, 1973; Gleitman & Wanner, 1984). However, research by Rushen Shi, Louann Gerken, and others has



argued that, even when children omit closed-class items from their speech, they perceive, represent, and actively use these items to learn about the grammatical organization of their language (Gerken & McIntosh, 1993; Shafer et al., 1998; Shi et al., 1999). Other research has suggested that there is a prosodic explanation for children's omissions. For example, laboratory experiments and studies of natural language acquisition have found that children are more likely to produce elements that bear stress, regardless of their meaning or grammatical function (Demuth & McCullough, 2009; Echols & Newport 1992; Gerken, Laudau & Remez, 1990; Pye, 1983), and Gerken (1991) has shown that young English speaking children are more likely to produce syllable sequences that form a trochaic (rather than iambic) pattern. Thus the absence of closed-class items from children's speech may reflect prosodic factors most relevant to production, rather than a failure to represent those items altogether. Children's omissions are therefore not evidence against the learning procedure we have proposed. However, more research is needed to understand what kinds of representations of closed-class items are necessary for the proposed learning procedure. Must children have complete, phonetically specified representations of a closed-class item in order to analyze it in this way? If so, this procedure will not be available until after these representations have developed, beginning in the second year of life for the highest-frequency items in the language (Shi et al., 2006).

Alternatively, children might be able to apply this learning procedure to an underspecified, collapsed category of closed-class items (Bloom, 1970; Echols & Newport, 1992; Mintz et al., 2002; Peters, 2001) or even to distinctive elements that they have just begun to learn (Marquis & Shi, 2012; Babineau et al., 2021), making the procedure more viable for the early stages of distributional analysis.

Another issue for future research concerns how this learning procedure operates in the acquisition of natural languages. At present, the evidence that learners analyze closed-class items in certain privileged ways is limited to the artificial and carefully designed experiments reported in this paper. To test our mathematical predictions most cleanly, these languages were unlike natural languages in a variety of ways: all categories other than X and Y were optional, there was only a single grammatical phrase (XY), and none of the words had meaning. We are in the process of asking whether learners privilege the same types of computations in the acquisition of languages that are more natural. This work promises to illuminate the kinds of natural language patterns that can be acquired and represented using these privileged computational mechanisms and can us help to understand whether these learning mechanisms may explain why these patterns have come to exist in languages of the world.

## References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324.
- Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, 40(5-6), 509–559.
- Babineau, M., de Carvalho, A., Trueswell, J., & Christophe, A. (2021). Familiar words can serve as a semantic seed for syntactic bootstrapping. *Developmental Science*, 24(1), e13010.
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: what level of representation? *Frontiers in Psychology*, 3, 451.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge: MIT Press.
- Braine, M. D. (1963). On learning the grammatical order of words. *Psychological Review*, 70, 323–348.
- Brown, R. (1973). *A first language: The early stages*. Harvard U. Press.
- Christophe, A., Guasti, T., Nespors, M., Dupoux, E., & Van Ooyen, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, 12(5-6), 585–612.
- Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: in support of language universals. *Cognition*, 139, 71–82.
- Culbertson, J., & Newport, E. L. (2017). Innovation of word order harmony across development. *Open Mind*, 1(2), 91–100.

- de Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and Function Words Cue the Acquisition of Word Meanings in 18-Month-Old Infants. *Psychological Science*, *30*(3), 319–332.
- Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language*, *36*(1), 173–200.
- Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology* (H.A. Ruger & C.E. Bussenius, Trans.). New York: Columbia University. (Originally published 1885.)
- Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition*, *2*(3), 189–220.
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*(44), 17897–17902.
- Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science*, *31*(2), 311–341.
- Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, *30*(4), 431–451.
- Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, *26*(2), 204–216.
- Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, *29*(3), 448–457.

- Gervain, J., Nespors, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive Psychology*, 57(1), 56–74.
- Gervain, J., Sebastián-Gallés, N., Díaz, B., Laka, I., Mazuka, R., Yamane, N., ... Mehler, J. (2013). Word frequency cues word order in adults: cross-linguistic evidence. *Frontiers in Psychology*, 4, 689.
- Gleitman, L. R., & Wanner, E. (1984). Richly specified input to language learning. In O. G. Selfridge, E. L. Rissland, & M. A. Arbib (Eds.), *Adaptive Control of Ill-Defined Systems* (pp. 227–249). Boston, MA: Springer US.
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), 341–353.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2), 151–195.
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: when learners change languages. *Cognitive Psychology*, 59(1), 30–66.
- Kramer, R. (2010). The Amharic definite marker and the syntax–morphology interface. *Syntax*, 13(3), 196–240.
- Marino, C., Bernard, C., & Gervain, J. (2020). Word Frequency Is a Cue to Lexical Category for 8-Month-Old Infants. *Current Biology: CB*. <https://doi.org/10.1016/j.cub.2020.01.070>
- Marquis, A., & Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition*, 122(1), 61–66.

- McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1–51.
- Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In *Action meets word: How children learn verbs* (pp. 31–63). Oxford University Press.
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393–424.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*(4), 498–550.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.
- Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535.
- Peters, A. M. (2001). Filler syllables: what is their status in emerging grammar? *Journal of Child Language*, *28*(1), 229–242; discussion 254–256, 269–274.
- Pye, C. (1983). Mayan telegraphese: Intonational determinants of inflectional development in Quiché Mayan. *Language*, *59*(3), 583–604.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, *66*(1), 30–54.
- Rizzi, L., & Cinque, G. (2016). Functional categories and syntactic theory. *Annual Review of Linguistics*, *2*(1), 139–163.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*(4), 493–515.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.
- Schuler, K. D., Yang, C., & Newport, E. L. (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2321–2326). Austin, TX: Cognitive Science Society.
- Schuler, K. D., Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). The effect of Zipfian frequency variations on category formation in adult artificial language learning. *Language Learning and Development*, *13*(4), 357–374.
- Shafer, V. L., Shucard, D. W., Shucard, J. L., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, *41*(4), 874–886.
- Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, *11*(3), 407–413.

- Shi, R., Marquis, A., & Gauthier, B. (2006). Segmentation and representation of function words in preverbal French-learning infants. In *Proceedings of the 30th Annual Boston University Conference on Language Development* (Vol. 2, pp. 549–560). Somerville, MA.
- Shi, R., & Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy*, *15*(5), 517–533.
- Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, *25*(1), 169–201.
- Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.
- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11–21.
- Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: the acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407.
- St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: the suffixing preference and grammatical categorization. *Cognitive Science*, *33*(7), 1317–1329.
- Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42.
- Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*(1), 71–86.



Zhang, Z., Shi, R., & Li, A. (2015). Grammatical categorization in Mandarin-Chinese-learning infants. *Language Acquisition*, 22(1), 104–115.