Privileged Computations for Closed-Class Items in Language Acquisition

Heidi R. Getz (hrg2@georgetown.edu) and Elissa L. Newport (eln10@georgetown.edu)
Center for Brain Plasticity and Recovery, Georgetown University Medical Center
Building D Suite 145, 4000 Reservoir Road NW
Washington, DC 20057 USA

Abstract

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 language learners' computations are 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 X could occur without Y. We created different versions of the language in order to ask whether learning depended on which of these categories was an open or closed class. In one condition, X was a closed class and Y was an open class; in a contrasting condition, X was an open class and Y was a closed class. Learning was significantly better with closed-class X, even though learners' exposure was otherwise identical. Additional experiments demonstrated that the perceptual distinctiveness of closed-class items drives learners to analyze them differently; and, crucially, that the primary determinant of learning is the mathematical relationship between closed- and open-class items and not their linear order. These results suggest that 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 may arise in part from this learning bias.

Keywords: language acquisition; statistical learning; computational mechanisms; morphosyntax; function words; closed-class items

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*.[1] 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 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 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 correlated distinctive phonological, prosodic and distributional properties such as short duration, light syllable structure, and high frequency (Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999), and children begin to represent these items long before producing them (Shafer, Shucard, Shucard, & Gerken, 1998; Shi, Werker, & Cutler, 2006). Early attention to closed-class items could facilitate other aspects of language acquisition. For example, since these items often occur at grammatically important parts of the sentence (e.g., phrase boundaries), focusing on them could help learners acquire grammatical structure. There is substantial empirical support for this idea, known as the Anchoring Hypothesis (Mintz, 2006; Morgan, Meier, & Newport, 1987; Valian & Coulson, 1988; Zhang, Shi, & Li, 2015).

However, it is not yet clear what computational mechanisms underlie learners' distributional analyses, once they have noticed closed-class items. The literature on statistical learning has not focused particularly on closed-class items; and only a few studies identify specific statistical computations that learners might draw on. These studies have revealed, 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 (Thompson & Newport 2007). 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, these items generally do not independently contribute semantic

[1] 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.

meaning; rather, they specify the grammatical properties 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.[2]

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 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 mang<u>es</u> pas?* ("You're not eating?), or non-finite as in *tu vas pas mang<u>er</u>?* ("You aren't 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 necessarily 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 proceed to learn their distribution. Because this distribution is asymmetrical—closed-class categories always predict but are not predicted by open-class categories—the computations that learners perform could 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.[3]

Here we explored the possibility that learners privilege 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 adult learners 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 learners' computations are biased: they identify patterns where closed-class items are predictive of open-class items more readily than the reverse. Additional experiments demonstrated that learners analyze closed-class items differently because they are perceptually distinctive (Experiment 2) and that learning outcomes are driven by the mathematical relationship between closed- and open-class items and not their linear order (Experiment 3). Together, the results suggest that learners analyze closed-class items in certain biased ways, searching preferentially for the kinds of patterns that exist in natural languages. In the Discussion we return to the question of why learners should be biased in this way. We do not mean to suggest that they know in advance about languages in particular, but rather that their computational biases may shape languages to be structured in this way.

Before proceeding, it is important to clarify a component of our experimental design. 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). Therefore X was the only category whose distribution with respect to other words was constrained; all other words in the language appeared and disappeared freely, which is unlike the more constrained sentence structure of natural languages. This

[2] 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.

[3] 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.

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, this would be motivation to explore how this bias affects the acquisition of more naturally structured languages—a line of work that is in progress.

Method

Participants

Three groups of sixteen adults from the Georgetown University community (age 18-28, mean=20.4) participated in this study. Two additional participants' data did not save due to an error.

Description of the miniature language

The design of the language used in all three experiments 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 only up to three categories could be omitted per sentence (i.e. sentences must each 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. 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 while in Experiment 3 X followed Y; this did not change the conditional relationship between the two terms.

In each experiment we created different versions of the language in order to ask whether learning this 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, so this contrast was defined by the number of words in each category and the phonological properties of those words. Each experiment had a condition where X was closed class and Y was open class (Closed X) and a contrasting condition where X was open class and Y was closed (Open X). Across experiments we varied the phonological properties of the closed-class item and the linear order of X and Y.

Experiment 1 In this experiment 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. Each open-class category contained three words that were a mixture of mono- and 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*. Thus the closed-class item *ka* predicted any of these three open-class items. In the contrasting Open X condition, X contained *lapal, tombur,* and *zup* and Y contained *ka*. Here the closed-class item *ka* is predicted by each of three open-class items.

Grammar: If X then Y*

AXYBC	XYBC	AYC	AY	YB
AXYB	ABC	XYC	AB	YC
AXYC	AXY	XYB	AC	BC
AYBC	AYB	YBC	XY	

Lexicon: Two conditions**

	A	X	Y	B	C
Closed X	flairb daffin glim	ka	lapal tombur zup	flugit mawg bleggin	clidam gentif spad
Open X	*(same)*	lapal tombur zup	ka	*(same)*	*(same)*

Example sentences

	Closed X	Open X
Experiment 1	ka_X $tombur_Y$ ★	$tombur_X$ ka_Y
Experiment 2	$daygin_X$ $tombur_Y$ ★	$tombur_X$ $daygin_Y$
Experiment 3	$tombur_Y$ ka_X ★	ka_Y $tombur_X$

*Sentence structures are shown for Experiments 1 and 2. In Experiment 3, sentences were the same except that X came after Y.
**Lexicon is shown for Experiments 1 and 3. In Experiment 2, the closed-class item was *daygin* instead of *ka*.

Figure 1: Design of the miniature languages used in Experiments 1-3. The critical feature of all languages is a one-way dependency between X and Y: every sentence with X also contained Y, but Y occurred without X. In Experiments 1 and 2, X came before Y (XY); in Experiment 3, X came after Y (YX). Each experiment had a condition where X was closed class and Y was open (Closed X) and a contrasting condition where X was open class and Y was closed (Open X). If learners are biased to analyze closed-class items as predictive, learning should always be better in the Closed X condition (marked with yellow stars).

Other than the specific lexical items in the X and Y categories, the two languages were identical. In both languages, 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 subjects' 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). Our manipulation did of course create different statistical patterns at the item level. In the Closed X condition, each of the three possible X-Y sequences had a transition probability of 0.33, whereas in the Open X condition each X-Y sequence had a transition probability of 1.0. Thus the item-level transition probabilities cued the XY unit more strongly in the Open X condition.

Because these languages are structurally 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 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.

Experiment 2 Part of the hypothesis is that the distinctiveness of closed-class items drives learners to analyze them differently. In Experiment 2 we tested this by making the closed-class item 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 is its high frequency. If distinctiveness of *ka* drove learning outcomes in Experiment 1, learning should be weakened in Experiment 2.

Experiment 3 In Experiments 1 and 2 the Closed X condition was superficially like English: the closed-class item came before the open-class item (Figure 1). English does also have closed/open dependencies where the closed-class item comes last (e.g., walk + s), but only in morphology. Therefore better learning for Closed X in Experiments 1 and 2 could be due to a preference in native speakers of English for syntactic phrases where frequent words come first (cf. Gervain et al., 2013). In Experiment 3 we changed the word order of the language so that Y preceded X. Now the Open X condition is superficially more like English (frequent word first), whereas the Closed X condition is superficially like Japanese (frequent word last). If learning outcomes in Experiments 1 and 2 reflect superficial word order biases, 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 rather than superficial linear order, results should be similar to those of Experiment 1: Closed X should learn "if X then Y" and Open X should fail.

Materials

We generated a 38-sentence exposure set by selecting two sentence types for each of the 19 structures. The sentence structures were always the same across conditions and experiments, but the actual sentence strings differed across conditions and experiments according to the lexical items in the X and Y categories and the linear order of X and Y. Sentence sound files were created by concatenating individually recorded words (spoken by a female native speaker of English) with 50 msec of intervening silence. The 38-sentence exposure set was presented 16 times as part of a 1-back task (see Procedure).

Procedure

Participants learned the language through a computer game. A robot "Bot" instructed participants to listen as an alien named Zooma practiced saying sentences in an alien language. After each sentence, participants pressed a button to indicate whether Zooma had just repeated herself. After exposure, participants began the test. Bot explained that Zooma would try to say each sentence two different ways, and the participants' job was to decide which one was better. The entire experiment lasted approximately 45 minutes.

Test

Learning was measured with a two-alternative forced-choice (2AFC) test. The structure of the test was identical across experiments. Specific test strings varied according to the vocabulary of the language. The target choice on each trial was always a grammatical complete sentence. The alternative was identical to the target except that either one word was changed, or the words were the same but in a different order. The test was designed to ask whether participants had acquired a very specific piece of knowledge: the precise 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, described below. Items with confounds (not scored) included four additional items testing XY constituency and 20 items testing the XY relationship within longer sentences. In addition, we included four items testing basic word order and six filler items in order to balance the frequency with which targets and foils for the critical XY trials appeared on the test. Results for these items are not described for space reasons, but they are generally consistent with the results here.

There were two trials for each of the critical test item types. One item type served as a constituency test: XY was compared to YB (Experiments 1,2) or XB (Experiment 3). Both choices are legal two-word sequences (in Experiments 1 and 2, both choices are also complete sentences). They have the same relative frequency in learners' input, and are medial bigrams in the basic sentence structure (AXYBC (Experiments 1,2) or AYXBC (Experiment 3)). However, X and Y are related grammatically whereas the elements in the foil sequence are not. A preference for XY would indicate that participants represent this grammatical relationship. A second item type (AY vs. AX) tested whether participants learned that X predicts Y, but not the reverse. In Experiments 1 and 2, both choices are legal two-word sequences and occurred in learners' input with the same relative frequency; 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. In Experiment 3, the foil was no longer a grammatical sequence in the language. Thus, it should be relatively easier in Experiment 3 than in Experiments 1 and 2 for both conditions to do well on these test items. Accuracy on the test was measured as choice of the target sequence.

Results

In Experiment 1 we asked whether learning of "if X then Y" would be different when X was a closed versus open class. Figure 2 illustrates that the answer is clearly yes. In the Closed X condition, learners chose the target item much more often than learners in the Open X condition (Closed X: 84%, Open X: 53%, $t(14) = 3.16$, $p = .007$). The grammatical coherence of XY as a unit was identical in these two conditions, since X always perfectly predicts Y. Yet learning outcomes differed across conditions, indicating that learners analyze closed-class items as predictive of open-class items more readily than the reverse.

Part of our hypothesis was that learners analyze closed-class items differently because these items are distinctive. In Experiment 2 we tested this by making the closed-class item less distinctive: here it was high frequency but phonologically like the open-class words in the language. Learning in the Closed X condition in this experiment was weaker than in Experiment 1 (Figure 2) and was no longer significantly better than the Open X condition ($t(14) = 1.94$, $p = .07$). This supports our hypothesis: closed-class items are analyzed differently because they are distinctive.

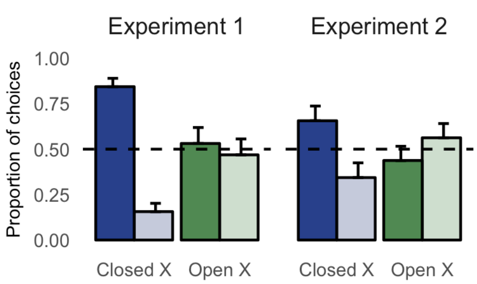

Figure 2: Choice of the target item (darker colors) or foil (lighter colors) on the 2AFC test in Experiments 1 and 2.

In Experiment 3 we asked: did the Closed X condition perform better because their language was superficially like English (the frequent item came first)? In this experiment Y came before X, but X still predicted Y. Thus the Open X condition was superficially like English. However, the Open X learners still struggled to learn "if X then Y" (59% correct, not significantly different from chance: $t(7)=2.05$, p = .08). In contrast, the Closed X condition continued to perform well above chance (72% correct, $t(7)=2.97$, p = .02), even though their language was superficially *unlike* English and more like an unfamiliar language, Japanese. These results demonstrate that the primary determinant of learning is the mathematical relationship between closed-class and open-class items and not their linear order.

The results just reported are collapsed across two types of items: the constituency test (YX vs. XB) and the predictive direction test (AY vs. AX). Based on these collapsed results, learning in the Open X condition appears to be slightly better than expected. Although learners were not significantly above chance, the difference was marginal, and accuracy was numerically higher than in Experiment 1 (59% vs. 53%). An analysis of results for the two different test item types provides some insight. In Experiments 1 and 2, results were equivalent across item types. However, the Open X condition in Experiment 3 showed a different pattern (Figure 3): learners passed the constituency test (88% correct), but were numerically *below* chance on the predictive direction test (33% correct). A 2-way mixed ANOVA over condition and trial type confirmed this impression statistically: there was no main effect of condition ($F(1) = 2.07$, $p = .17$), but there was a significant main effect of test item type ($F(1) = 4.77$, $p = .047$), and—importantly—a significant interaction between condition and test item type ($F(1) = 7.45$, $p = .02$), driven by a preference in the Open X condition for the ungrammatical sequence *AX over AY.

Why would the Open X condition perform so poorly on the AY/*AX items? Based on the raw statistical properties of learners' input, these items should be easy: *AX is not a complete sentence or even a legal sequence, whereas AY is both. In fact, an explanation for these results is provided by our hypothesis: that learners attend to (or search for) some statistical patterns over others, prioritizing patterns in which closed-class items are predictive. Such a bias would lead learners in the Open X condition to initially analyze their closed-class item Y as predictive of another item. Statistically, the item that Y best predicts is X (the probability that a sentence contains X, given that it contains Y, is .53). Thus, a preference for *AX could reflect an incorrect hypothesis that the conditional relationship between X and Y is reversed ("if Y then X"). This generalization is not consistent with learners' input, but it is consistent with the patterning of closed-class items in natural languages.

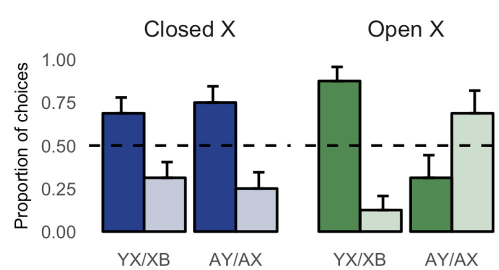

Figure 3: Choice of the target item (darker colors) or foil (lighter colors) on the two item types of the 2AFC test in Experiment 3. Participants in Closed X still learned, even though the linear order was opposite English. In contrast, participants in Open X succeeded on the constituency test (YX/XB) but not the predictive direction test (AY/AX), apparently having incorrectly analyzed Y as predictive of X.

Discussion

In three experiments, we showed that adults easily acquire a grammatical dependency "if X then Y" when X is a closed-class, but fail to acquire the same dependency when X is an open class (Experiment 1). Successful learning of "if X then Y" is facilitated by the distinctive perceptual properties and high frequency of the closed-class item (Experiment 2). Importantly, the primary determinant of learning is the mathematical relationship between closed-class and open-class items and not their linear order (Experiment 3). These results suggest that learners privilege computations in which closed-class items are predictive of open classes—the same computations that are most relevant for natural language dependencies.

We emphasize that, within each experiment, the Closed X and Open X conditions had exactly the same statistical evidence for the rule "if X then Y": class X always perfectly predicted class Y. Furthermore, learners in contrasting conditions were always exposed to the same number of lexical items, sentence structures, and sentence types. To acquire the XY rule, learners needed to compute exactly three transitional probabilities, and contrasting conditions within each experiment always contained the same linear direction of the X-Y relationship (forward for Experiments 1 and 2, backward for Experiment 3). Despite this mathematical equivalence, learning was always better when X was a closed class. If participants were computing statistics over items rather than classes, the results are even more striking: in that case, participants learned three low-probability dependencies with a predictive closed-class item (*ka*→*lapal*, *ka*→*tombur*, *ka*→*zup*) more easily than three high-probability dependencies with a predictive open-class item (*lapal*→*ka*, *tombur*→*ka*, *zup*→*ka*).

These results indicate that—whether learners computed statistics over classes or items—their distributional analyses are biased. Rather than tracking all possible pairwise transitional probabilities involving a closed-class item, learners apparently analyze closed-class items asymmetrically, more easily learning patterns in which a closed-class item is predictive *of* another element than patterns in which it is predicted *by* another element.

Conclusion

The original idea of the Anchoring Hypothesis (Valian & Coulson, 1988) was that, because closed-class items tend to occur at grammatically important points in the sentence, focusing on them could help learners acquire grammatical structure. Our results add a computational component to this approach. Our hypothesis is that, 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.

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 modern syntactic theory as well. The results of our experiments suggest that human languages may acquire this type of structure at least in part as a consequence of computational biases in the human language learner. This account is appealing because, if correct, it would explain the privileged role of closed-class items in human linguistic representations without positing that these representations are innate. However, it is important to note that in all three experiments, learners' preferred conditional relationship had the same abstract structure (though not always the same linear order) as the closed/open dependencies in all natural languages, including English. It is difficult to rule out the possibility that learning was affected by participants' experience with this abstract property of natural languages; even infants have experience with closed-class items (cf. Shi et al., 1998). Studies of learning in a non-linguistic domain could be informative (cf. Saffran, Johnson, Aslin, & Newport, 1999).

Our results raise several other important questions. First, what about closed-class items that behave differently? Our claim is that learners analyze closed-class items as predictive of open-class items, and that this approach is useful because it matches the abstract structure of grammatical dependencies in natural languages. However, there are exceptions to this pattern. For example, pronouns like *him* do not depend on open-class items the same way that articles do. Interestingly, pronouns are also special in other ways (Chomsky, 1980). The proposed computation could be useful not only for discovering predictive dependencies, but also—when this analysis fails to uncover a dependency—for flagging elements with a more complex grammatical distribution. Second, we must also ask whether this computational bias is present in children, who are the real natural language learners. Our results in ongoing work with child participants suggest that they do share this bias. This in turn raises a puzzle: if children organize their languages around closed-class items, why do they not produce these words in their own speech for several years? The available evidence suggests that children do indeed process closed-class items early, despite omitting them in production (Gerken et al., 1990; Shi et al., 2006; Zhang et al., 2015). Future work is required to understand the discrepancy between what children represent and what they initially produce. Finally, we need to test our predictions on materials that are more like natural languages than what we have studied here. In order to test our computational predictions most cleanly, the languages in these experiments were unlike natural languages in several ways: all of the categories other than X and Y were optional, there was only a single grammatical phrase (XY), and none of the words had any meaning. We are in the process of testing whether learners privilege the same types of computations in the acquisition of miniature languages that are more natural. If so, we can ask what kinds of natural language patterns can be acquired and represented using these privileged computational mechanisms, and to what extent these learning mechanisms explain why these patterns come to exist in languages of the world.

Acknowledgements

This research was supported by NIH grant HD037082 to E. Newport, by the Feldstein Veron Fund for Cognitive Science, and by Georgetown University Center for Brain Plasticity and Recovery.

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.

Chomsky, N. (1980). On binding. *Linguistic Inquiry, 11*(1), 1–46.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, *26*(2), 204–216.

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.

Kramer, R. (2010). The Amharic definite marker and the syntax–morphology interface. *Syntax*, *13*(3), 196–240.

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). New York: Oxford University Press.

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.

Rizzi, L., & Cinque, G. (2016). Functional categories and syntactic theory. *Annual Review of Linguistics*, *2*(1), 139–163.

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.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.

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., 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.

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.