The Role of Prior Learning in Biasing Generalization in Artificial Language Learning

Jill Lany (lany@u.arizona.edu)
Department of Psychology, The University of Arizona, Tucson, AZ 85721-0068

Rebecca Gómez (gomez@u.arizona.edu)
Department of Psychology, The University of Arizona, Tucson, AZ 85721-0068

Abstract

Because unbiased learners are unlikely to arrive at the appropriate generalizations of their language (Gold, 1967), accounts of acquisition must examine the nature of learning biases. One form of bias is learners' prior learning experience. Adult participants familiarized with a category-induction language learned a language with the same underlying structure but novel vocabulary much more rapidly than naïve learners (Lany, Gómez, & Gerken, 2004). In the present experiments, we extend our investigations of prior learning experience by manipulating whether learners were initially exposed to fully- or partially-cued structure. Generalization is hindered by prior exposure to fully-cued structure, but enhanced by prior exposure to structure that is partially-cued. The results are important for understanding of the role of prior experience in constraining language acquisition.

Introduction

The syntax of natural languages is highly complex. Without corrective feedback, unconstrained learners are unlikely to converge on the grammar of their linguistic community (Gold, 1967). However, language learners regularly manage to discover the underlying patterns of their language while ignoring irrelevant structure, suggesting that they are constrained. Recent work in artificial language learning has begun to investigate forms of learning constraints, or ways in which learning processes might be guided.

Saffran (2002) demonstrated modality constraints on learning an artificial language by exposing adult learners to a phrase-structure language in which the presence of one item in a phrase predicted the presence of another item (Language P), or to a phrase-structure language in which predictive relationships were absent (Language N). Participants were presented with either an auditory or visual version of Language P or N. Learning of Language P was better when presentation was auditory. Visual language learners acquired the predictive and non-predictive languages equally well, suggesting that learners' sensitivity to patterns and relationships varies as a function of modality.

Recent work by Gómez (2002; Gómez, Welch, & Lany, 2003) speaks to another way learning may be guided. Gómez' studies suggest that learners selectively tune in to regularities by seeking out the most reliable structure in their input. While learners are highly sensitive to conditional probabilities of adjacent elements (e.g., Saffran, Aslin, & Newport, 1996), Gómez demonstrated that when conditional probabilities between adjacent elements are low or unreliable, learners attend to relationships between non-adjacent elements. These findings indicate that learners are biased to take the statistical reliability of a structure into account. Importantly, languages make use of various cues highlighting relevant structure, not just statistical ones[1].

Saffran's (2002) work suggests that learners are biased before they even begin to learn language. Gómez's work (2002; Gómez et al., 2003) suggests that statistical characteristics of the input itself can bias learning. Research also suggests learners can be biased by prior experience.

Saffran and Thiessen (2003) provided evidence that once infants form phonological generalizations based on regularities in their input, those generalizations influence how they parse new speech materials. Lany, Gómez, and Gerken (2004) demonstrated that generalization occurs at the abstract level of syntax-like structure. In these experiments, adults were exposed to a language consisting of categories of words, with restrictions on how categories could combine (see Braine, 1987; Frigo & MacDonald, 1998; Gerken, Wilson, & Lewis, in press; Gómez & Lakusta, in press). The language is composed of words belonging to the categories *a*, *b*, *X*, and *Y*. As in natural languages, the rules

[1] For example, languages are rich with prosodic cues to syntactic structure, and learners are sensitive to such cues beginning in early infancy (see Jusczyk, 1997 for an overview).

involve relationships between word categories. Specifically, the language has restrictions on how categories of different types can be combined within a string, such that *a* elements are paired with *X* elements and *b* elements with *Y*s, but not vice versa. (See Figure 1.)

	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	$\mathbf{X_5}$
$\mathbf{a_1}$	a_1X_1	a_1X_2	a_1X_3	a_1X_4	a_1X_5
$\mathbf{a_2}$	a_2X_1	a_2X_2	a_2X_3	??	a_2X_5

	$\mathbf{Y_1}$	$\mathbf{Y_2}$	$\mathbf{Y_3}$	$\mathbf{Y_4}$	$\mathbf{Y_5}$
$\mathbf{b_1}$	b_1Y_1	b_1Y_2	b_1Y_3	b_1Y_4	b_1Y_5
$\mathbf{b_2}$	b_2Y_1	b_2Y_2	b_2Y_3	??	b_2Y_5

Figure 1. A typical *aX bY* paradigm. Learners are exposed to a subset of the grammatical pairings of markers and content-words then are tested for generalization to the withheld (??) pairings.

Learners exposed to 18 or 6 minutes of the *aX bY* language successfully acquired the language in the former but not the latter condition. However, learners exposed to 18 minutes of one language, and then transferred to a second language with the same underlying pattern, but none of the same words, learned the pattern with just 6 minutes of exposure.

The findings of Lany et al. (2004) demonstrate that learners do not remain the same over the course of acquisition. Rather, the learning process changes them, constraining the ways they perceive and learn about subsequent input. Thus, prior experience represents an additional constraint enabling learners to successfully acquire language.

Using this procedure, we can begin to investigate other ways learning might be facilitated by prior experience. In doing so, it is important to explore the extent to which this process might be useful in natural language acquisition.

In English, consistency in head direction results in a co-occurrence relationship between determiners and nouns, and also between auxiliaries and verbs. These relationships both involve a functional element preceding a lexical one, and restrictions on co-occurrences of categories of functional and lexical elements (similar to the *aX bY* structure used by Lany et al., 2004). Importantly, learners only truly acquire categories, and their co-occurrence restrictions, when there are cues indicating category membership (Frigo & McDonald, 1998; see also Gerken et al., in press). However, cues in natural language are variable, differing with strength according to category. For example, a corpus analysis (Lany et al., 2004) showed that in infant-directed speech, a greater proportion of nouns are marked with morpho-syntactic cues than verbs – e.g. nouns were fully cued by a determiner and a plural or diminutive ending 20% of the time and partially cued by either a determiner or an ending 60% of the time. Verbs were fully cued both by an auxiliary and inflectional ending 1% of the time and by one or the other 20% of the time. Thus languages contain different degrees of fully- and partially-cued structure.

In this study, we asked whether learners with prior exposure to a well-cued pattern have an advantage over naïve learners in acquiring a version of that pattern in which the cues highlighting relevant structure are diminished. In Experiment 1, we exposed a control group to 18 minutes of a partially-cued *aX bY* language, in which only 60% of the *X* and *Y* words had cues to category membership. We exposed an experimental group to 18 minutes of a fully-cued *aX bY* language, in which 100% of the *X* and *Y* words had cues to category membership, and then transferred them to 18 minutes of the partially-cued *aX bY* language. Interestingly, learners with prior exposure to a fully-cued language did not subsequently learn a partially cued language better than the control group. However, in Experiment 2, learners with prior exposure to a partially-cued language learned a second partially-cued language better than naïve learners. Why might this be the case? Learners initially exposed to the fully-cued language may have learned the perfectly predictive surface regularities resulting from the underlying structure, as opposed to the category relationships. These surface regularities were probabilistic in the partially-cued language, and thus a focus on this aspect of structure would not result in successful learning at transfer. In other words, perhaps learners of the fully-cued language were hindered by their focus on the surface regularities of the strings. Learners of the partially-cued language were not led to rely exclusively on a cue that would later be less reliable. We speculate on factors facilitating this group's transfer in the discussion.

Experiment 1

Method

Participants Ninety-five University of Arizona undergraduate students participated for course credit, forty-eight in the experimental condition.

Materials Learners in these experiments were exposed to category-induction languages of the form *aX bY*. We constructed both a fully-cued and a partially-cued *aX bY* language. The fully-cued language had two versions which shared the same underlying structure, but had none of the same words (two versions were necessary to test transfer). The vocabulary of each version of the language consisted of two *a* elements (*ong* and *rud* in Version A, and *ush* and *dak* in Version B) and two *b* elements (*alt* and *pel* in Version A, and *erd* and *vot* in Version B). The vocabulary of each version also consisted of six bisyllabic *X* words and six bisyllabic *Y* words. The *X* words all ended with the same syllable ("-ul" in Version A and "-it" in Version B). Similarly, all *Y* words shared the same final syllable ("-ee" in Version A and "-oo" in Version B. Additionally, each version had two variants, or grammars. The two grammars of a version were composed of the same set of words, but differed in how they were combined – strings from Grammar 1 (G1) took the form *aX bY* and strings from Grammar 2 (G2) took the form *aY bX*. For example, strings from Version A, G1were *ong bivul* and *erd suffee*, and strings from G2 were *ong suffee* and *erd bivul*. In Version B, G1 strings were *ush zamit* and *alt wifoo*, and G2 strings were *ush wifoo* and *alt zamit*. The fully-cued versions of the language consisted of 24 possible strings (12 *aX* strings and 12 *bY* strings), however, 4 strings (2 *aX* and 2 *bY* strings) were withheld from familiarization to be presented at test, so the familiarization set consisted of 20 strings.

The test materials for the fully-cued languages consisted of 16 strings, half grammatical and half ungrammatical. Four grammatical strings had been withheld during familiarization, four strings had been presented during training, and eight strings were ungrammatical strings (these were from the unheard language). Strings that were grammatical for one group of participants were ungrammatical for the other.

The materials for the partially-cued languages were the same as those of the fully-cued languages, with the exception that only 60% (four of six) of the *X* and *Y* words had endings cueing their category membership. Uncued words were bisyllabic, and each had a distinct ending (i.e. an ending that was not present on any of the other words, cued or uncued). Examples are *jeeloff*, *skyjer*, *bowda*, and *pefto*. These uncued words replaced four of the cued words from the fully-cued version of the language. The partially-cued versions of the language consisted of 24 possible strings (12 *aX* strings and 12 *bY* strings), however, 8 strings (4 *aX* and 4 *bY* strings) were withheld from familiarization to be presented at test, so the familiarization set consisted of 16 strings.

The test materials for the partially-cued languages consisted of 32 strings, half grammatical and half ungrammatical. Eight grammatical strings had been withheld from familiarization (four were cued and four were uncued). Eight strings had been presented during familiarization (four were cued, four were uncued). There were also 16 strings from the grammar of the unheard language.

The process underlying successful learning of this language is twofold (Braine, 1987; Frigo & MacDonald, 1998). Learners must first discover that there are different categories of words, which requires that words from different categories be differentiable based on their semantic or phonological characteristics. Once learners are sensitive to the categories, they can then learn that there are restrictions on how categories co-occur. Learners with knowledge of co-occurrence restrictions can generalize to novel combinations that respect these restrictions. When cues to category membership are present, generalizations can be accomplished through attention to the pairing of *a*s and *b*s with the endings of the *X* and *Y* words. Additionally, learners exposed to the partially cued language, can generalize to novel combinations involving uncued words by noting that if an *X*-word is paired with a particular *a* word, it can co-occur with other *a* words, but not with *b* words.

Procedure There were eight conditions in this experiment, resulting from the between-subjects manipulations of familiarization type (Transfer vs. Control), version (Version A vs. Version B), grammar (Grammar 1 vs. Grammar 2), Aside from instructions at the start of the familiarization phase, which were delivered by the experimenter, the entire experiment was conducted on a Hewlett Packard Brio PC running SuperLab 2.01 software.

In the Transfer condition, participants listened over headphones to 18 blocks (approximately 18 minutes) of randomly ordered strings from their fully-cued training language, and then answered two iterations of 16 test

questions, each in a different random order and separated from each other by a brief pause. Using the "Y" and "N" keys on their keyboard, participants made yes/no judgments on the grammaticality of each string. They then repeated this familiarization and test procedure for a partially-cued version of the language with novel vocabulary. In the Control condition, participants were familiarized with 18 blocks of the partially-cued language before test.

For the fully-cued language, participants sensitive to the *aX bY* structure should endorse grammatical test strings, including those withheld during training, more often than they endorse ungrammatical ones. Similarly, for participants familiarized with a partially-cued language, sensitivity to the *aX bY* structure would be indicated by higher endorsement rates to withheld grammatical strings (both cued and uncued) than to ungrammatical ones.

Results and discussion Preliminary analyses indicated that there were no differences in learning as a function of language version or grammar, so we collapsed across these variables. Mean endorsement rates to test strings are found in Table 1.

We tested whether participants in the Transfer group learned the partially-cued language better than Control participants. We did a three-way mixed ANOVA, with a between-participant factor of familiarization type (Transfer vs. Control), and the within-participant factors of test string familiarity (heard vs. unheard), and test string cues (cued vs. uncued). The dependant measure was the difference in endorsement rates to grammatical test strings and their ungrammatical counterparts.

There was a main effect of test string familiarity, $F(1, 92) = 88.89$, $p < .001$, with the difference in endorsement rates to heard grammatical strings and ungrammatical ones ($M = .27$, $SE = .026$) more than to unheard ones ($M = .05$, $SE = .022$). There were no other main effects or interactions. Thus, both Transfer and Control participants endorsed heard grammatical test strings more often than unheard grammatical strings, but the Transfer group did not perform better than the control group. Neither group showed differences in endorsement rates to *unheard* grammatical and ungrammatical strings, ts ≤ 1.82, ps $\geq .076$. Thus there was no generalization to unheard items.

These findings suggest that Transfer learners did not benefit from their prior exposure to a fully-cued *aX bY* language. One explanation is that learners cannot acquire an *aX bY* pattern with only partial cues, regardless of their prior experience. However, given that other studies provide evidence of learning partially cued *aX bY* structure (e.g. Frigo & McDonald, 1998), this explanation seems unlikely. An alternative explanation is that learners exposed to a fully-cued language focused only on the surface relationship between the *a* and *b* elements and endings on the *X* and *Y* words, essentially learning a co-occurrence relationship between the first word in the string and the ending on the second. If this were the case, they would learn only the surface regularities resulting from the underlying structure as opposed to the category relationships. Experiencing a perfectly predictive relationship between the first word and the ending of the second may have led Transfer learners to tune-in to this aspect of the partially-cued language as opposed to the abstract structure.

We next tested whether learners would transfer from a partially-cued *aX bY* language, by exposing them to a partially-cued *aX bY* language before transferring them to another version of this partially-cued language. Because the initial language does not have a perfect correspondence between the initial word and ending of the final word of a string, learners would not likely focus solely on it, and thus might perform differently than the Experiment 1 learners at transfer.

Table 1: Endorsement Rates to Test Items with Standard Errors in Parentheses

	Grammatical Heard		Grammatical Unheard		Ungrammatical	
Expt 1	Cued	Uncued	Cued	Uncued	Cued	Uncued
Transfer	.76 (.028)	.77 (.031)	.54 (.029)	.60 (.028)	.50 (.031)	.55 (.032)
Control	.80 (.020)	.75 (.026)	.62 (.025)	.56 (.031)	.56 (.029)	.59 (.026)
Expt 2						
Transfer-2	.83 (.028)	.80 (.026)	.59 (.036)	.54 (.031)	.48 (.033)	.48 (.032)

Experiment 2

Method

Participants Forty-eight University of Arizona undergraduates participated for course credit.

Materials The language materials were the two partially-cued versions of the language used in Experiment 1.

Procedure The procedure for Experiment 2 participants was the same as for the Transfer group in Experiment 1, such that participants were familiarized and tested on one version of the language in Phase 1, and then transferred to the other version in Phase 2. Half of the participants were exposed to Version A of the partially-cued language and then Version B, and half to Version B and then A. This group is referred to as the Transfer-2 group.

Results and discussion We wanted to determine whether participants in the Transfer-2 group learned the partially-cued language they heard in Phase 2 better than the control participants from Experiment 1. Thus, we did a three-way mixed ANOVA, with the between-participant factor of familiarization type (Transfer-2 vs. Control), and the within-participant factors of test string familiarity (heard vs. unheard), and test string cues (cued vs. uncued). The dependant measure was the difference in endorsement rates to grammatical and ungrammatical test strings. Mean endorsement rates to test items can be found in Table 1.

We found that the difference in the Transfer-2 learners' endorsement rates to grammatical vs. ungrammatical test items ($M = .21$, $SE = .030$) was higher than that of the control group from Experiment 1($M = .11$, $SE = .031$), $F\,(1, 93) = 5.69$, $p = .019$. There was an effect of test string familiarity, with the difference in endorsement rates to heard grammatical test strings and ungrammatical ones ($M = .27$, $SE = .026$) greater than the difference between unheard grammatical test items and ungrammatical ones ($M = .05$, $SE = .022$), $F\,(1, 93) = 102.45$, $p < .001$. There was also an effect of test string cues. The difference in endorsement rates to grammatical test strings with cues and ungrammatical strings ($M = .20$, $SE = .026$) was significantly greater than the difference to grammatical test strings without cues and ungrammatical strings ($M = .13$, $SE = .023$), $F\,(1, 93) = 6.50$, $p = .012$. There were no interactions between any of the three variables, suggesting that learning in the Transfer-2 group was generally better than the Control group.

Paired-sample *t* tests comparing endorsement rates for each of the four types of grammatical strings and ungrammatical ones indicate that the Transfer-2 group learned the underlying structure of the language (see Table 1 for means and standard errors). Endorsement rates to grammatical heard test strings, cued and uncued, were higher than those to ungrammatical ones, ts $(47) \geq 6.36$, ps $\leq .001$. Critically, endorsement rates to *unheard* grammatical strings with cues were higher than those to ungrammatical strings, $t\,(47) = 2.10$, $p = .04$. (Recall that control subjects in Experiment 1 did not show such learning). Endorsement rates to grammatical unheard strings without cues did not differ from those to ungrammatical ones, $t\,(47) = 1.68$, $p = .1$.

In sum, learners with prior exposure to a partially-cued language subsequently learn a new version of such a language better than naïve learners. Thus, these findings shed light on how learners might acquire patterns in the absence of robust cues typically necessary for successful learning.

General discussion

In this set of experiments, we demonstrated that what learners can acquire from their input changes as they gain experience with a particular type of structure. Our results suggest that learners exposed to a fully-cued category-induction language become sensitive to a phonological pattern in the form of a perfect correspondence between *a* and *b* words and the endings on the *X* and *Y* words. Because they have not become sensitive to the underlying category relationships, only to the surface correlates of this pattern, their learning of a partially-cued language in which this relationship is probabilistic is not enhanced relative to controls. However, exposure to a partially-cued language facilitates subsequent acquisition of another partially-cued language. What is the basis for generalization in these learners? Recall from Experiment 1 that control participants exposed to a partially-cued language do not generalize to new strings. Tranfer-2 learners, whose initial learning phase was identical to controls', are also unlikely to have learned the *aX bY* structure. While neither the Transfer nor the Transfer-2 group appeared to successfully

acquire the structure of their initial training language, unlike learners transferred from a fully-cued language, Transfer-2 learners were not led to rely on a cue that would later be disrupted. While we cannot precisely determine what aspect of their exposure to the partially-cued language facilitated generalization, it is clear that strong sensitivity to the underlying category relationships does not drive the effect. Generalization may be driven by sensitivity to the underlying pattern learners acquire in their initial exposure. Generalization may also be influenced by other similarities between languages. In these experiments, strings from both versions of the language were composed of two words, the first monosyllabic and the second bisyllabic, and most *X*s and *Y*s in both versions of the language had distinctive features in the form of endings on the words. If some or all of these similarities were absent, transfer of structure might be less likely to occur, thus raising important questions about constraints on transfer. We might test this by transferring learners to languages in which some of these similarities are removed.

These results add to our previous work (Lany et al., 2004) by providing information about how prior learning experience can interfere with generalization (Exp. 1) or enhance it (Exp. 2). We plan to extend this work by asking whether prior exposure to one of the partial-structure languages facilitates acquisition of a language in which the cues are even further diminished. This would be analogous to asking whether the higher incidence of partially cued noun phrases found in English child-directed speech could help learners acquire verb phrases (in which category structure is less reliably marked). This manipulation should also provide information about how prior learning affects the flexibility of later learning.

In conclusion, learners' prior experience can be instrumental in shaping what they acquire from their input. Learning biases of other sorts have also been proposed, such as constraints on the kinds of computations likely to be performed based on the modality of the input (Saffran, 2002), and the tendency to acquire the most reliable or statistically predominant structure from indeterminate input (Gómez, 2002; Gómez et al., 2003). Prior experience is yet another source of bias constraining language learners. Understanding the scope of these biases will contribute importantly to our theories of language acquisition.

Acknowledgements

This research was supported by NIH R01 HD42170-01 to RLG.

References

Braine, M.D.S. (1987). What is learned in acquiring word classes – A step toward an acquisition theory. In B. McWhinney (ed.), *Mechanisms of Language Acquisition.* Hillsdale, NJ: Erlbaum.

Frigo L., & MacDonald, J. (1998). Properties of phonologicasl markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39,* 448-457.

Gerken, L., Wilson, R., & Lewis, W. (in press). Seventeen-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language.*

Gold, E.M. (1967). Language identification in the limit. *Information and Control, 10,* 447-474.

Gómez, R.L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431-436.

Gómez, R.L., & Lakusta, L. (in press). A first step in form-based category abstraction in 12-month-old infants. *Developmental Science.*

Gómez, R.L., Welch, K., & Lany, J. (2003). Statistical determinants of learning. Paper presented at BUCLD, Boston, MA.

Jusczyk, P.W. (1997). *The Discovery of Spoken Language.* Cambridge, MA: The MIT Press.

Lany, J., Gómez, R.L, & Gerken, L. (2004). Generalization to parallel structure in language acquisition. Manuscript submitted for publication.

Saffran, J.R. (2002). Constraints on statistical learning. *Journal of Memory and Language, 47,* 172-196.

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

Saffran, J., & Thiessen, E. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39,* 484-494.