Journal of Memory and Language 65 (2011) 1–14

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Balancing generalization and lexical conservatism: An artificial language study with child learners

Elizabeth Wonnacott *

The University of Oxford, UK

A R T I C L E I N F O

Article history:
Received 17 September 2010
revision received 22 February 2011
Available online 1 April 2011

Keywords:
Language acquisition
Statistical learning
Artificial language learning
Over-generalization
Baker's Paradox

A B S T R A C T

Successful language acquisition involves generalization, but learners must balance this against the acquisition of lexical constraints. Such learning has been considered problematic for theories of acquisition: if learners generalize abstract patterns to new words, how do they learn lexically-based exceptions? One approach claims that learners use distributional statistics to make inferences about when generalization is appropriate, a hypothesis which has recently received support from Artificial Language Learning experiments with adult learners (Wonnacott, Newport, & Tanenhaus, 2008). Since adult and child language learning may be different (Hudson Kam & Newport, 2005), it is essential to extend these results to child learners. In the current work, four groups of children (6 years) were each exposed to one of four semi-artificial languages. The results demonstrate that children are sensitive to linguistic distributions at and above the level of particular lexical items, and that these statistics influence the balance between generalization and lexical conservatism. The data are in line with an approach which models generalization as rational inference and in particular with the predictions of the domain general hierarchical Bayesian model developed in Kemp, Perfors & Tenenbaum, 2006. This suggests that such models have relevance for theories of language acquisition.

Introduction

Successful language acquisition requires an ability to generalize, but learners must balance this against the acquisition of exceptions. For example, English native speakers regularly combine adjectives and nouns in novel ways, yet there are some arbitrary restrictions on usage: *strong winds* may be *high winds*, but *strong breezes* are not? *high breezes.* These preferences do not arise from any obvious semantic constraints – rather they appear to rely on knowledge that is arbitrary and lexically-based. Another example is verb sub-categorization preferences. For example, a number of English verbs can occur in both of two near synonymous dative structures: the prepositional-dative, e.g. *I told the story to Ben*, and the double-object-dative, e.g. *I told Ben the story*, yet others are restricted to occur only the prepositional-dative, as in *I explained the story to him, *I explained him the story.* Again, these lexical exceptions are not obviously explained by semantic constraints.

Children (at least from age 3, see Tomasello, 2000) are able to generalize and use words in new ways, as seen both in overgeneralization errors (*Don't say me that!*) and the usage of nonce words in new constructions in experiments (e.g. Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989). How then do they learn that certain usages are impermissible? This quandary, known as *Baker's Paradox* (Baker, 1979), has been considered central to theories of language acquisition. One solution attempts to show that apparent exceptions actually arise from more general regularities based on the semantic or phonological properties of words (Pinker, 1989). However, although such factors influence generalization in adults and older children (e.g. Gropen et al., 1989), young children may not be sensitive to the requisite conditioning criteria (Brooks & Tomasello, 1999;

* Address: Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK. Fax: +44 (0) 1865 310447.
E-mail address: elizabeth.wonnacott@psy.ox.ac.uk

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.03.001

Ambridge, Pine, Rowland, & Young, 2008). Their role in development is thus unclear. Moreover, various researchers have disputed the claim that these cues can fully determine verb-syntax (e.g. Braine & Brooks, 1995), suggesting that learners must be capable of learning arbitrary, lexically-specified exceptions.

An alternative approach, originating with Braine (1971), suggests that learners are sensitive to lexically based distributional statistics and use this information to make inferences about when generalization is and is not appropriate. This concurs with approaches to language acquisition that emphasize the role of statistical learning processes (Elman, 1990; Rumelhart & McClelland, 1986; Saffran, Aslin & Newport, 1996). There is evidence that children and adults are more likely to allow novel generalizations with low frequency lexical items. For example, children are more likely to over-generalize with low frequency than high-frequency verbs, judging "*He came me to school*" to be worse than "*He arrived me to school*" (Ambridge et al., 2007; Brooks, Tomasello, Dodson, & Lewis, 1999; Theakston, 2004). This has been explained in terms of *Entrenchment* (Braine & Brooks, 1995) or *Statistical Pre-emption* (Goldberg, 2005)[1]: frequently encountering verbs with alternative constructions leads to reluctance to generalize to a new construction.

In addition to the frequency of lexical items, higher-level statistics may also affect generalization. Goldberg (2005) argued that high-frequency verb argument structures such as the transitive are more likely to be generalized to new verbs. Corroborating evidence comes from the sentence processing literature. When reading or listening to language, we make predictions about upcoming sentence structure which concur with the verb's distributional history (e.g. Trueswell, Tanenhaus, & Kello, 1993), but may also be influenced by "higher-level" biases – for example to interpret post-verbal nouns as direct objects even for intransitive verbs (Mitchell, 1987). This latter effect may be due to the greater frequency of the transitive structure across the language (Juliano & Tanenhaus, 1994). Such findings about language *processing* suggest that language *learning* involves accumulating statistics at both a lexically specific and more generalized level.

A recent study by Wonnacott et al. (2008) provided direct evidence that adult learners can use both lexically specific and higher-level statistics in constraining generalization. In order to establish that effects were driven by distributional patterns at and above the lexical level, rather than semantic or phonological motivations, an Artificial Language Learning methodology was used, i.e. participants were exposed to experimenter-created miniature languages and tested to see when they generalized. Specifically, the miniature languages incorporated two competing transitive structures; generalization occurred when a participant used a verb with a structure with which it did not occur in the input. In addition to individual verb frequency, two 'higher-level' statistical factors were found to influence the usage of a verb in a novel construction (a) the frequency of the structure across the language: more generalization with a higher frequency structure (b) the distribution of verb types across the language: if *most* verbs across their input language had occurred in both of the two structures (so called *alternating* verbs), learners were more likely to generalize verbs from one construction to the other. These effects were particularly clear with what Wonnacott et al. called *minimal-exposure* verbs. These were verbs which occurred in only one of the two structures and with very few exposures (four). Importantly, they were presented to learners only *after* they had been previously exposed to a large amount of language input involving other verbs. Wonnacott et al. asked whether learners would restrict their usage of these verbs to the structure in which it had been encountered (lexical conservatism), or extend it to the other structure (generalization). From the perspective of individual lexical frequency, four exposures is a very small sample, learners might therefore be expected to ignore this verb-specific input and generalize. In fact, learners' treatment of these verbs depended upon the input to which they had been previously exposed: participants previously exposed to a language where all verbs occurred in just one structure (dubbed the Lexicalist language), showed strong conservatism and little generalization; in contrast, learners exposed to a language where all verbs occurred in both structures (dubbed the Generalist language), generalized those verbs to both structures, particularly generalizing the structure that was of higher frequency across the language.

Wonnacott et al. argued that their learners were taking a *rational* approach to determining when to generalize from minimal evidence, drawing on a theoretical framework provided by Bayesian approaches to cognition. This was formalized by Perfors, Tenenbaum, and Wonnacott (2010) who demonstrated that the data are in line with the predictions of a hierarchical Bayesian model (henceforth HBM). This domain general model had been developed by Kemp, Perfors, and Tenenbaum (2007), who applied it to a distinct set of cognitive learning problems (for example, the problem of acquiring the "shape bias" in word learning), yet it could predict the behavior of the adult artificial language learners. Critically, the model is characterized by an ability to track statistical distributions at multiple levels of abstraction, and to make inferences about the extent to which these levels provide a good indicator of future behavior. This is achieved via the formation of "over-hypotheses" about a particular data set. For example, when it was trained on the Lexicalist language from Wonnacott et al. (2008), the model formed an "over-hypothesis" to the effect that the usage of constructions was highly consistent for particular verbs, whereas in the Generalist language it formed the over-hypothesis that verb identify and construction usage were unrelated. These over-hypotheses led to the model showing the same difference in the learning of minimal-exposure verbs as human learners, i.e. more learning of the lexical constraints in the Lexicalist than Generalist languages. The model also mimicked human learners in showing greater generalization with the more frequent of the two constructions, due to the fact that it tracked their distribution across the whole language.

[1] The entrenchment and statistical pre-emption hypotheses are subtly different since the latter assumes that novel verb-structure pairings, or other generalizations, are only blocked by encountering near synonymous alternatives. This difference is beyond the scope of the current work since the structures which are generalized in our experiments carry no semantics.

What is the relevance of such higher-level learning for theories of language acquisition? One difficulty in interpreting the findings of Wonnacott et al. (2008) is that all participants were adults. Although there are similarities in adult and child language learning, there are also important differences (Newport, 1990). One possibility is that the rational usage of both lexically-based and higher-level statistics might result from access to deliberate learning strategies which would be unavailable to children. Thus to establish the relevance of this type of learning for theories of acquisition, it is important to extend the results to children. This is the central purpose of the current study.

Although artificial language learning has a long history with adults (e.g. Braine, 1963) and there are also several studies with infants (e.g. Gerken, 2006; Gomez, 2002; Saffran, Newport, & Aslin, 1996), studies with older (verbal) children are relatively few, and none have explored the balance between generalization and lexical restrictions (though see Brooks, Braine, Catalano, Brody, & Sudhalter, 1993). Most relevant is a set of statistical learning experiments which looked at the process of *regularization*,[2] where learners create a systematic linguistic system out of unsystematic input (Austin, Newport, & Wonnacott, 2006; Hudson Kam & Newport, 2005; Hudson Kam & Newport, 2009). Specifically, these experiments explored how children (5–6 year olds) and adults learned artificial languages where nouns were followed by words referred to as "determiners", although they were generally meaningless. The general finding was that, given languages where the usage of particular determiners was variable and probabilistic, children would *regularize* – for example, consistently using a single particle across productions, or making particle usage dependent on sentence position. In contrast, given the same input, adults tended to reproduce the probabilistic usage in their input. In line with this finding, Wonnacott et al. (2008) found that when adults were exposed to the Generalist language described above – i.e. a language where the verbs in the input occurred probabilistically with each of the two transitive structures – they continued to use the structures probabilistically, matching their probabilities of occurrence in the input.

The current experiments explored how children balance lexically-based learning and generalization using an artificial language paradigm similar to that used by Hudson Kam and Newport. As a secondary consideration, where appropriate, the data were also examined for evidence of regularization. Experiment 1 explored whether children show the same sensitivity to higher-level statistics as the adults in Wonnacott et al. (2008). Experiment 2 explored whether lexically-based learning was affected by lexical frequency.

Experiment 1

Two artificial languages were created, where the critical relationships were between nouns and meaningless words which are here called "particles". In order to focus on children's learning of the syntactic relationships, the languages involved a mixture of familiar English nouns and novel function words. Following Wonnacott et al. (2008), the languages were dubbed the Lexicalist and Generalist languages. Although the critical relationships were between nouns and particles, rather than verbs and constructions, these languages encompassed the critical properties of the equivalent languages in Wonnacott et al. (2008). In the Lexicalist language each of the nouns occurred consistently with just one particle (*particle1-only* and *particle2-only* nouns). In the Generalist language, each noun occurred with both of the two particles (*alternating* nouns). In addition, in both of the languages, one of the particles was more frequent. Particle usage was tested in a production test following exposure. As in Wonnacott et al. (2008), most critical were minimal-exposure test items. These involved two nouns which were not in the initial input but were each presented in four sentences during the testing session, and always with the same particle (i.e. one noun four times with particle1: *minimal-exposure-particle1-only* noun; one noun four times with particle2: *minimal-exposure-particle2-only* noun). The prediction was that we would see greater learning of the association between these nouns and particles (i.e. fewer productions with the unattested particle) after exposure to the Lexicalist language than after exposure to the Generalist language, since only the Lexicalist input provides evidence that particle usage is lexically-conditioned. A second prediction was that, when generalization did occur, we would see greater extension of the more frequent particle1.

Following Wonnacott et al. (2008), we also elicited productions with familiar nouns (from the exposure sets) and entirely-novel nouns (occurring only in testing). The prediction for familiar nouns was that particle productions would reflect the patterns of usage in the input (although Hudson Kam & Newport, 2005 suggests the possibility of regularization in the Generalist language). The prediction for entirely-novel nouns was that when children generalized particle usage to these nouns there would be more generalization of the more frequent particle1.

Method

Participants

Forty two children (5;3–7;1 years) were recruited from Year 1 classrooms from schools in Oxfordshire. 21 were exposed to the Generalist language and 21 to the Lexicalist language (mean age 6;0 in each condition). All were monolingual native English speakers. Each child was tested individually in two separate sessions of approximately 20 min held on two consecutive days.

Language conditions

Children were exposed to a different set of sentences depending upon their assigned language condition. Each set consisted of 16 sentences that were repeated three times across the two experimental sessions. Every sentence described a picture of two cartoon animals and was three words long. The form of the sentence was *moop NOUN PARTICLE,* where *moop* is a novel verb meaning

[2] Note that "regularization" is distinct from "generalization" which in this paper specifically refers to the situation where a child produces an unattested noun particle bigram. In the Hudson Kam & Newport studies all noun–particle combinations had actually occurred in the input and so "generalization" in this sense was not possible.

Table 1
Noun types and their exposure in the Lexicalist and Generalist conditions.

	Noun type	Number of nouns of this type	Number of times each noun encountered with particle 1	Number of times each noun encountered with particle2
Lexicalist-language exposure	*Particle1-only*	3	12	0
	Particle 2-only	1	0	12
Generalist-language exposure	*Alternating*	4	9	3
Additional nouns presenting in testing (both language conditions)	*Minimal-exposure particle1*	1	4	0
	Minimal-exposure particle 2	1	0	4

"THERE-ARE-TWO", NOUN is the relevant English label for the animal (e.g. "pig"), and PARTICLE was either particle1 or particle2, specifically the new words *dow* and *tay* (with assignment as particle1 and particle2 counterbalanced across participants) and carried no semantics. Four nouns were used in the input in each of the two language conditions: in the Lexicalist condition there were three *particle1-only* nouns and one *particle2-only* noun; in the Generalist condition there were four *alternating* nouns. These nouns were drawn from the following set: *giraffe*, *pig*, *rabbit*, *cow*, *cat*, *dog*, *mouse*, *crocodile* (the remaining four nouns were reserved for use as minimal-exposure and entirely-novel nouns in testing). Half of the participants in each condition received a fixed assignment of particular nouns to noun types, and for the remaining half the assignment was randomized on a participant to participant basis.[3] In both conditions, each of the four nouns was equally frequent in the input set. For the Lexicalist condition one exposure set comprised: four sentences with each of the three *particle1-only* nouns (always followed by particle1), four sentences with the single *particle2-only* noun (always followed by particle2). For the Generalist condition one exposure set comprised: four sentences with each of four *alternating* nouns (each noun followed by particle1 in three sentences and particle2 in one sentence). The exposure for each of the different noun types in each condition is summarized in Table 1.

Note that in both languages, particle1 is three times more frequent than particle2, and the frequency of the particles is matched across the two languages.

Procedure

A two day procedure was used, with tests at the end of the second session.

Day 1

Introduction and noun practice. Children were told that they were going to meet some animals who "say things differently from us" and that they would learn how to say some things the way the animals did. They then viewed pictures of the eight animals which would occur in training and testing, and practiced their "names" (e.g. *pig/rabbit*). Where children used alternative names for the animals than those in the training set (e.g. "bunny" for rabbit) we would encourage usage of the correct name ("he likes to be called rabbit – can you say rabbit?").

Sentence exposure. Children were presented with the set of 16 sentences in their language condition twice through, taking a short break in the middle. Sentences were presented in random order. Each trial consisted of viewing a picture of two identical cartoon animals, hearing a sentence and copying it aloud (e.g. seeing TWO-PIGS, and hearing and repeating *moop pig tay*). If they made errors in copying the sentences (e.g. saying the wrong animal name) they were asked to listen and repeat again.

Day 2

Noun review. Children viewed each picture again and named the animal aloud.

Sentence exposure. Children were again exposed to the 16 sentences in their language condition, exactly as on Day1.

Familiar nouns test. Children were presented with each of the pictures used in exposure and asked to provide their own sentences. To get them started, they were given the first word (i.e. "*moop*..."). There were eight test trials (four nouns ∗ two trials each) presented in random order. If the child produced the wrong noun (e.g. *pig* instead of *cow*) he/she was encouraged to have another attempt (*"is that pig?" "can you try again?"*). The children would then generally correct themselves, though these items were excluded from data analysis.[4] Children's productions were transcribed by the experimenter and recorded for later reference.

Exposure and testing of minimal-exposure nouns. Using the same procedure as before, children were presented with a set of eight sentences involving new nouns not in their exposure set. There were two nouns, *minimal-exposure-particle1-only* and *minimal-exposure-particle2-only*, which each occurred four times and always with the same

[3] Whether participants received the fixed versus random vocabulary assignment was initially included as a factor in all analyses in order to ensure that the original "fixed" vocabulary assignment did not accidentally aid (or hinder) learning. It is excluded from the reported results since, as expected, there were no significant main effects or interactions for this factor.

[4] In pilot experiments some children began ignoring the content of the pictures half way through testing – presumably they assumed that if they had not been corrected this was permissible.

particle. For example, if *minimal-exposure-particle1-only* was *dog* and *minimal-exposure-particle2-only* was *cat*, children would encounter *moop dog tay* (TWO-DOGS) ∗ four and *moop cat dow* (TWO-CATS) ∗ four. Immediately afterwards, they were asked to provide their own sentences for these nouns, using the same procedure as previously, with four test items for each of the two nouns.

Entirely-novel nouns test. Children were asked to provide sentences for two new nouns using the same procedure as in the previous test. The pictures were of animals seen in noun practice but not occurring in the input or as minimal-exposure nouns.

Results

Data for familiar, minimal-exposure, and entirely-novel nouns are analyzed separately. In each case, since we are interested in children's choice of particle usage, trials were excluded in which children did not produce a "correct" (*moop*) *NOUN PARTICLE* (irrespective of sub-categorization) restrictions). (Errors included using an incorrect noun – even if subsequently self-corrected; producing English or random utterances (e.g. "*moop carrots*"); producing *moop NOUN* and something incorrect in place of the particle (e.g. *"moop cat rumarama", "moop crocodile snapping", "moop cat forgotten"*); not producing any particle[5]; refusing to respond). Children were not penalized for omitting to repeat the initial "moop"). The percentage of included sentences provides a baseline measure of learning and is reported in each case.

Familiar nouns

Baseline

87% of productions in the Generalist condition and 97% of productions in the Lexicalist condition had the correct (*moop*) *NOUN PARTICLE* form. An independent *t*-test revealed these figures to be significantly different ($t(24.681) = 2.24$, $p = .034$, *df* adjusted due to unequal variances; all *t*-tests two-tailed), indicating that children find the Generalist language somewhat harder to learn.

Choice of particle

The proportion of (*moop*) *NOUN PARTICLE* productions with each of the two particles for each noun type is shown in Fig. 1. In the Lexicalist condition, children show little generalization to the unattested particles: they produced 98% particle1 with the *particle1-only* nouns (2% particle2), 81% particle2 (19% particle1) with the *particle2-only* noun. This shows strong learning of the lexically-based patterns. A paired *t*-test confirmed that particle1 was used significantly more often with the *particle1-only* nouns than *particle2-only* nouns ($t(20) = 10.524$, $p < .01$). In the Generalist condition, both particles are used with the alternating nouns but, as in the input, particle1 is used approximately three times more often with particle1 (72% particle1 usage). A one sample *t*-test showed this to be significantly different from chance (i.e. 50%) usage of each particle ($t(21) = 4.672$, $p < .01$).

Given the results of Hudson Kam and Newport (2005) I also looked for *regularization* in the Generalist condition – i.e. are individual children using particles systematically? A child was categorized as a "regularizer" if, for every noun where they produced two (*moop*) *NOUN PARTICLE* sentences, they produced the same particle in both sentences. Note that this captures both regularization where the child uses the same particle across all nouns, and regularization on a noun by noun basis. According to this definition, seven children regularized (six of these used the higher frequency *particle2* across all nouns and one regularized noun-by-noun – see Appendix A for the full pattern of productions for each child). What is the probability of seeing this number of "regularizers" by chance? If a child was randomly producing the two particles with a 75:25 particle1:particle2 ratio, the probability of producing the same particle across the two productions with a given noun is 0.625. In order to work out the probability that (at least) a given number of children will regularize, it is necessary to first group them according to the number of nouns for which they produced multiple (i.e. two) (*moop) NOUN PARTICLE* productions, since this affects their probability of "chance" regularization. First, nine children produced two (*moop*) *NOUN PARTICLE* utterances for all four nouns and thus each had a $p = .153$ probability of regularizing (either across nouns or on a noun by noun basis). Three of these children actually regularized, and the probability of seeing (at least) this number regularizing by chance was calculated to be $p = .146$ (all calculations use the Binomial Equation – details are given in Appendix B). Second, eight children produced two (*moop*) *NOUN PARTICLE* utterances for exactly three nouns and thus each had a $p = .244$ probability of regularizing. Three of these children actually regularized, and the probability of seeing (at least) this number regularizing by chance was calculated to be $p = .307$. Third, three children produced two (*moop*) *NOUN PARTICLE* utterances for exactly two nouns and thus had a $p = .391$ probability of regularizing. One of these children regularized, and the probability of seeing (at least) this number regularizing by chance was calculated to be $p = .773$. (The remaining child did not produce two (*moop*) *NOUN PARTICLE* utterances for any noun and so had no opportunity to show regularization). The combined probability of seeing at least the number of regularizers we did in each group is $p = .035$, suggesting above chance regularization (see Appendix B for full details of calculations).

Minimal-exposure nouns

Baseline

90% of productions in the Generalist condition and 98% of productions in the Lexicalist condition had the correct (*moop*) *NOUN PARTICLE* form. These means are significantly different ($t(24.373) = 2.689$, $p = .013$; df adjusted for inequality of variance), again indicating that children have more difficulty in the Generalist condition.

[5] Bare "(*moop*) *NOUN*" productions, i.e. with no particle and nothing in its place, were rare. They accounted for only 10% of excluded sentences and less than 1% of total productions.

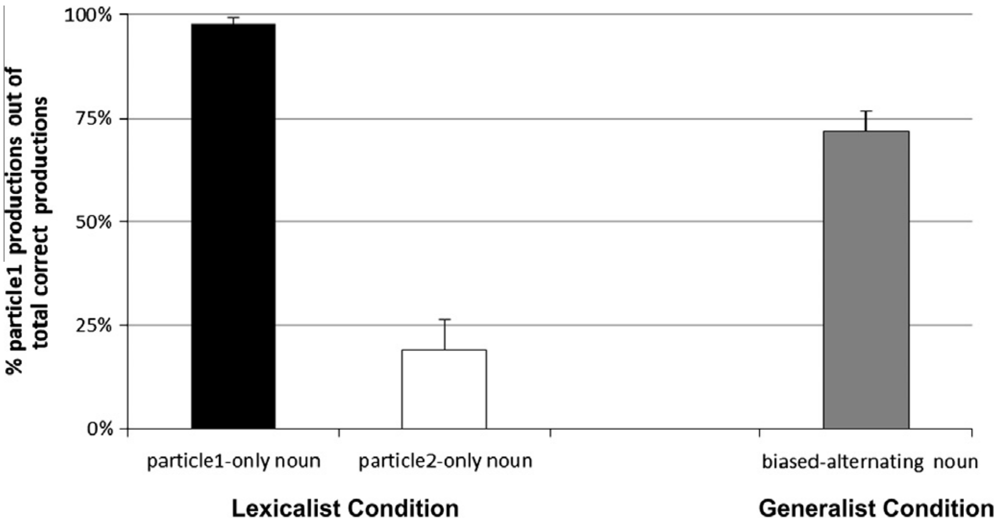

Fig. 1. Experiment 1: familiar nouns in the Lexicalist and Generalist conditions. (Note: as all correct sentences have the form (*moop*) *NOUN PARTICLE*, 0% particle1 productions indicates 100% particle2 productions.)

Choice of particle

Fig. 2 shows children's particle usage. As predicted, there is evidence that children have learned the associations between the two nouns and their attested particles in both conditions (i.e. more particle1 productions with the *minimal-exposure-particle1-only* noun, more particle2 productions with the *minimal-exposure-particle2-only* noun), but the difference is greater in the Lexicalist condition (despite the fact that exposure to the relevant noun–particle bigrams was matched across conditions). These effects were shown to be significant in an ANOVA with *Percentage-Particle1-Productions* as the dependent measure, *Noun-Type* as a within-subject factor with two levels (*minimal-exposure-particle1-only* and *minimal-exposure-particle2-only*) and *Language* as a between-subjects factor with two levels (Lexicalist vs. Generalist). There was no main effect of *Language* ($F(1, 40) = 1.197$, $p = .28$), but a main effect of *Noun-Type* ($F(1, 40) = 43.101$, $p < .001$) and a significant interaction between *Noun-Type* and *Language* ($F(1, 40) = 7.709$, $p = .008$).

Is generalization more common with particle1 (which was three times more frequent in both languages)? One indicator of this is the total particle1 usage for (*moop*) *NOUN PARTICLE* productions with the two minimal-exposure nouns. If the particles were equally generalized this should approximate 50%. In fact, in the Generalist condition, 59% of these productions used particle1. This is significantly greater than 50% usage ($t(20) = 2.426$, $p = .025$), indicating more generalization with the more frequent particle. In the Lexicalist condition there was 52% usage of particle1, which was not significantly greater than 50% ($t(20) = 0.382$, $p = .707$). This suggests that the statistical prevalence of particle1 in the input influences the treatment of minimal-exposure nouns for learners of the Generalist language but not for learners of the Lexicalist language.

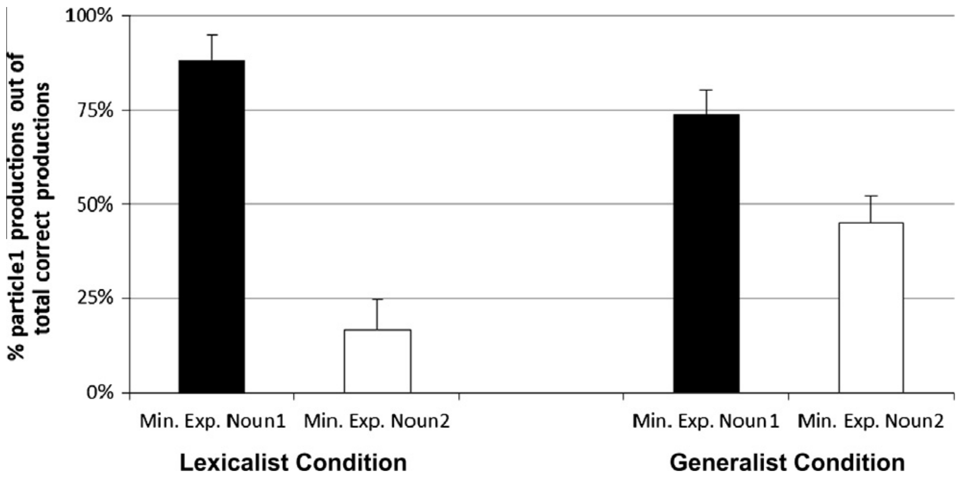

Fig. 2. Experiment 1: minimal-exposure nouns in the Lexicalist and Generalist conditions. (Note: as all correct sentences have the form (*moop*) *NOUN PARTICLE*, 0% particle1 productions indicates 100% particle2 productions.)

Entirely-novel nouns

Baseline

85% of productions in the Generalist condition and 83% of productions in the Lexicalist condition had the correct (*moop*) *NOUN PARTICLE* form. These means are not significantly different ($t(40) = 0.178$, $p = .860$). Note that all such productions are generalizations, since the nouns have not previously been encountered with either particle. Since these baseline scores are numerically lower than for familiar nouns, I also ran an ANOVA with *Percentage-Correct* as the dependent measure, *Noun-Type* a between-subjects condition with two levels (familiar versus novel) and *Language* a within-subjects condition with two levels (Lexicalist versus Generalist). There were no significant effects (main effect *Noun-Type:* $F(1, 40) = 2.250$, $p = .141$; *Language*: $F(1, 40) = 0.58$, $p = .449$; interaction $F(1, 40) = 1.148$, $p = .290$). This suggests that children are not finding it significantly harder to produce correct sentences with novel nouns than they did with familiar nouns.

Choice of particle

The percentage of (*moop*) *NOUN PARTICLE* productions with each of the two particles is shown in Fig. 3. (Note that 2/21 children in the Generalist condition and 2/21 children in the Lexicalist condition produced no correct productions with these nouns and thus do not contribute towards these data). In both conditions, we see that particle1 usage is greater than particle2 usage (Lexicalist condition: 75% particle1 usage, significantly greater than chance (i.e. 50%), $t(18) = 3.479$, $p = .003$; Generalist condition 60% particle1 usage, marginally greater than chance (i.e. 50%), $t(18) = 1.817$, $p = .086$). A paired *t*-test did not reveal a significant difference between the two conditions ($t(36) = 1.665$, $p = .105$). This suggests that children are influenced by the fact that, in both conditions, particle1 is more frequent across the nouns of the input.

Although previous work with children has not explored regularization with novel nouns (though see Wonnacott & Newport, 2005), I also looked for evidence of regularization here. Again regularization was defined as consistently using one particle per noun – either the same particle across both nouns or a different one with each. In the Generalist condition 5/19 possible children regularized (three produced particle1 across all productions, two produced 100% particle1 with one noun and 100% particle2 with the other). In the Lexicalist condition 13/18 possible children regularized (8 produced particle1 across all productions, 1 produced particle2 across all productions, 4 produced 100% particle1 on one noun and 100% particle2 on the other; note that children had to produce multiple (*moop*) *NOUN PARTICLE* sentences with at least one noun in order to have any opportunity to show regularization, the full pattern of productions is given in Appendix A). The probability of seeing this number of children regularize by chance probability matching was calculated in a similar manner as for familiar nouns in the Generalist language (see Appendix B for details). The results were: Generalist condition $p = .021$, Lexicalist condition $p < .000$. In addition, a Chi-Square test was used to compare the number of children regularizing in each condition. I restricted this comparison to include only children who produced two (*moop*) *NOUN PARTICLE* sentences for all four nouns (although including all children yields equivalent results). There were 14 such children in the Generalist condition, and 3 of them regularized, and 11 such children in the Lexicalist condition, and nine of them regularized. The comparison (i.e. 3/14 versus 9/11) revealed a significant difference (*Chi-squared* = 6.744, *df* = 1, $p = .009$), indicating greater regularization in the Lexicalist condition.

Discussion

In Experiment 1 two groups of 6 year olds were each exposed to an artificial language in which nouns occurring within sentences were obligatorily followed by one of two meaningless particles. They were then tested to see whether and how they produced particles when producing their own sentences with various noun types. The central question was how children would balance the usage of

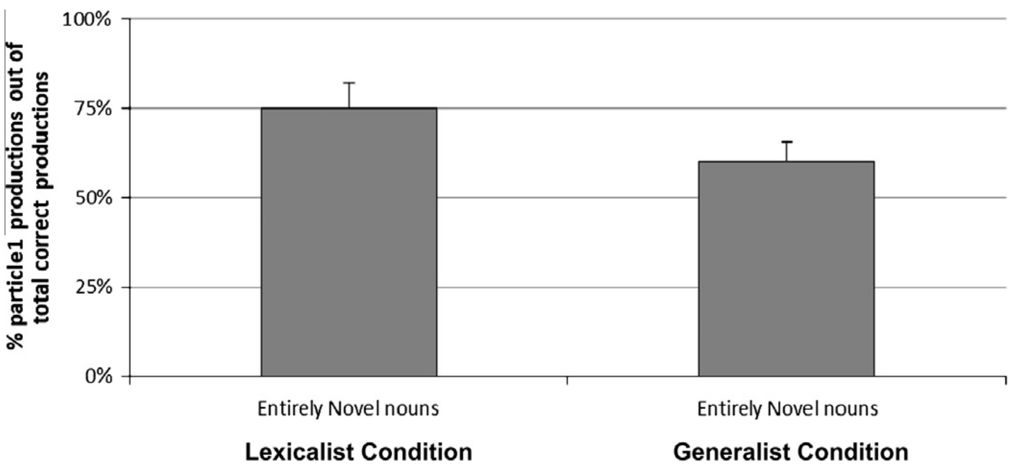

Fig. 3. Experiment 1: entirely-novel nouns in the Lexicalist and Generalist conditions. (Note: as all correct sentences have the form (*moop*) *NOUN PARTICLE*, 0% particle1 productions indicates 100% particle2 productions.)

lexically-based information (i.e. using nouns with the particles with which they occurred in the input) against generalization (using unattested noun–particle combinations). The general hypothesis was that children would show evidence of both lexically-based learning and generalization, but that the balance between the two would be influenced by the overall statistical properties of their input language. Children's ability to replicate lexically-based patterns is first seen with familiar noun test items. As predicted, children's productions indicated that they had learned how particles should be used with the nouns in their input: learners of the Lexicalist language showed almost perfect learning, virtually never using the familiar *particle1-only* and *particle2-only* nouns with the unattested particles. Meanwhile learners of the Generalist language closely mimicked the production probabilities of two particles for the biased alternating nouns in their input. Thus children showed an ability to learn lexically-based patterns of usage which are both absolute and probabilistic (although, interestingly, children produced more "incorrect" productions in the Generalist condition, suggesting that unconditioned probabilistic patterns of usage may be harder to learn than absolute constraints).

Lexically-based learning is also seen with the minimal-exposure nouns. Although these nouns had only been encountered four times, children in both language conditions continued to primarily use those nouns with the attested particle. However we also saw evidence of generalization– i.e. productions with the unattested particle – and, critically, there was significantly more of this after exposure to the Generalist language. The exposure for these specific noun–particle bigrams was matched across the two conditions, as was the frequency of each the two particles across the nouns in the input. Thus the difference in children's productions indicates that they are influenced by their past experience of the relationships between nouns and particles: they are much more likely to use a new noun with a particle it did not occur with if their previous experience suggests that particle usage is not lexically restricted (i.e. after exposure to the Generalist language). This result mirrors the findings with minimal-exposure verbs in Wonnacott et al. (2008), and accords with the predictions of the HBM from Kemp et al., 2007 and Perfors et al., 2010. (I return to discuss how this model accounts for the data in the General Discussion).

In the Generalist condition, the minimal-exposure nouns also demonstrate the effects of another higher-level statistical factor: the noun-independent frequency of the particle across the language, i.e. we see more generalization with particle1 than particle2. Interestingly, although the frequency of the two particles was matched in the Generalist and Lexicalist conditions, this statistic plays no role in the treatment of the same minimal-exposure nouns in the Lexicalist condition. This is explained by the children's much stronger reliance on lexical patterns in that language. In other words, when they generalize children are influenced by the frequency of the two particles across the language, but generalization is balanced against the past behavior of the specific nouns and this wins out in the Lexicalist Condition. The treatment of entirely-novel nouns is also in line with this explanation. Recall that here children have *no* experience with the specific nouns in question. Still, in both conditions, children produce correct (*moop*) *NOUN PARTICLE* productions the majority of the time (and not significantly less than with familiar nouns) – showing a clear ability to generalize. Critically, in both languages, they are more likely to use particle1 than particle2, showing an influence of the language-wide prevalence of the two particles.

A secondary question was whether there was any evidence of regularization in these experiments, as has been witnessed in previous artificial language learning experiments with child learners. We first looked at the familiar nouns in the Generalist condition – the learning condition most similar to previous experiments. We found that more children regularized than would be predicted if each child was randomly probability matching their usage of particles in (*moop*) *NOUN PARTICLE* sentences. This fits with the results of previous studies with children, although the amount of regularization was less than previous experiments might lead us to expect – a point to which I return in the final discussion. We also looked at entirely-novel nouns in both conditions. In each condition more children regularized than would be predicted by chance, however we saw significantly more regularizers in the Lexicalist condition than in the Generalist condition. Since there are no other differences between the conditions, this difference must derive from the different expectations children have formed with regards to the consistency of particle usage in the input languages, i.e. that in the Lexicalist language particle usage should not vary for individual nouns. Note that this bias could lead either to regularization of the same particle across both nouns, or regularization of a different particle per noun since both patterns are lexically consistent (though if children have also learned that *particle1-only* nouns are three times more frequent they should be most likely to regularize that particle across both nouns – which is indeed the most frequent pattern). Note also that if previous exposure to lexically conditioned particle usage affects regularization, the fact that *entirely-novel* nouns are tested *after* the *minimal-exposure* nouns exposure, might possibly explain why these nouns show as much regularization as they do in the Generalist condition, i.e. if that exposure leads to a weak bias for lexical consistency. Regardless, the difference in the extent of regularization in the two conditions provides further evidence that children have learned something about the higher-level relationships between nouns and particles in the input languages. This concurs with the findings with minimal-exposure nouns, and is also in line with the HBM presented in Perfors et al. (2010), a point to which I again return in the General Discussion.

The results of the current experiment demonstrate that the extent to which children rely on lexically-based patterns can be affected by statistical factors above the level of individual lexical items. This was particularly clear for minimal-exposure nouns, where children in the Generalist condition showed more generalization due to their previous experience of alternating nouns in the input. While this is in line with the findings of Wonnacott et al. (2008), and the HBM (Perfors et al., 2010), this result raises the question of how children are able to acquire natural languages where some lexical items alternate in their usage of some linguistic structures but others are restricted to just one structure (e.g. *give/send/throw* can occur with both the prepositional and double object dative, whilst *donate/transport* are restricted to occur only with one). If encountering alternating lexical items encourages generalization, can learners nevertheless show sufficient learning of lexical restrictions?

Wonnacott et al. (2008) found that adults learning "mixed" languages, which included both alternating and restricted verbs, *could* learn the restrictions on non-alternating verbs, but this factor was modulated by the frequency of the verb. This is also predicted by the Entrenchment and Statistical Pre-emption hypotheses and the HBM. Experiment 2 used the noun–particle paradigm to explore whether children can also learn lexical restrictions in "mixed" languages, and whether such learning is greater when the relevant nouns (and hence noun–particle bigrams) are of higher frequency.

Experiment 2

Two new semi-artificial languages were created, dubbed MixedLow and MixedHigh. Both contained one *particle1-only* noun, one *particle2-only* noun and two *alternating* nouns. The input sets for the two conditions were identical except that the sentences with *particle1-only* and *particle2-only* nouns were three times more frequent in MixedHigh. The prediction was that we would see learning of the lexical patterns in both conditions, but stronger learning of the lexical restrictions - i.e. less usage of *particle1-only* and *particle2-only* nouns with the unattested particle – in the MixedHigh condition.

Method

Participants

29 children (5:4 to 7:0 years) were recruited from Year 1 classrooms from schools in Oxfordshire. All were monolingual native English speakers. Fifteen were assigned to the MixedLow condition (mean age 6;2) and 14 to MixedHigh (mean age 6:1).

Language conditions

There were two language conditions, MixedHigh and MixedLow. In MixedLow the exposure set consisted of 16 sentences: four with a *particle1-only* noun, four with a *particle2-only* noun, four with each of two alternating nouns (two per particle). In MixedHigh the exposure set consisted of 32 sentences: 12 with a *particle1-only* noun, 12 with a *particle2-only* noun and four with each of two alternating nouns (two with each particle). The exposure for each of the different noun types in each condition is summarized in Table 2.

Procedure

The procedure was identical to that used in Experiment 1 except that the experiment ended at the end of the familiar nouns test.

Results

Baseline. 89.0% of productions in MixedLow and 88.1% in MixedHigh had the correct (*moop*) *NOUN PARTICLE* form. An independent *t*-test revealed that these figures were not significantly different ($t(27) = 0.141$, $p = .889$), suggesting equal learning of the basic word order pattern in the two languages.

Choice of particle. We then examined children's choice of particle usage for (*moop*) *NOUN particle* productions. The relevant data are shown in Fig. 4.

We first considered children's usage of the particles with the two *alternating* nouns. If children's usage of particles matches the input, they should use each particle around 50% of the time. We saw 54% particle1 usage in MixedLow and 38% particle1 usage in MixedHigh. One sample t-tests suggested that neither mean differed from 50% usage (MixedLow: $t(14) = 0.562$, $p = .582$; MixedHigh: $t(13) = 1.319$, $p = .210$). The two means also did not significantly differ from each other ($t(27) = 1.370$, $p = .182$),

Most critical are productions with the *particle1-only* and *particle2-only* nouns. As can be seen in Fig. 4, in both conditions, children produce particle1 more often with the *particle1-only* noun and particle2 more often with the *particle2-only* noun, however this difference is more pronounced in the MixedHigh condition. An ANOVA was run with *Percentage-Particle1-Productions* as the dependent measure, *Noun-Type* as a within-subject factor with two levels (particle1-only versus particle2-only) and *Language* as a between-subjects factor with two levels (MixedLow versus MixedHigh[6]). This revealed no main effect of *Language* ($F(1, 25) = 0.184$, $p = .672$), but a significant main effect of *Noun-Type* ($F(1, 25) = 30.404$, $p < .001$), and a significant interaction between *Noun-Type* and *Language* ($F(1, 25) = 5.811$, $p = .024$).

Although our main question in this experiment concerned the frequency manipulation, it is also interesting to compare the learning of *particle1-only* and *particle2-only* nouns in the MixedLow condition with their counterparts in the Lexicalist condition from Experiment 1. These nouns (and thus the relevant noun–particle bigrams) were matched in frequency across the two conditions, so any difference in how they were learned must reflect other higher-level properties of the language. In the Lexicalist condition, 98% of correct productions with *particle1-only* nouns used particle1, and 19% of correct productions with *particle2-only* nouns used particle1 (see Fig. 1) – a 79% difference, compared with 29% in the current MixedLow condition. To compare these differences, an ANOVA was conducted with *Percentage-Particle1-Productions* as the dependent measure, *Noun-Type* as a within-subject factor with two levels (particle1-only versus particle2-only) and *Language* as a between subjects factor with two levels (Lexicalist versus MixedLow). This showed no significant main effect of *Language* ($F(1, 33) = 0.356$, $p = .555$), but a main effect of *Noun-Type* ($F(1, 33) = 46.469$, $p = .000$) and a significant interaction ($F(1, 33) = 8.882$, $p = .005$). This indicates

[6] Two children in the MixedLow condition produced no correct productions with one of the two nouns and thus contribute no data.

Table 2
Noun types and their exposure in the MixedLow and MixedHigh conditions.

	Noun type	Number of nouns of this type	Number of times each noun encountered with particle1	Number of times encountered with particle2
Mixed language low	*Particle1-only* (*low frequency*)	1	12	0
	Particle2-only (*low frequency*)	1	0	12
	Alternating	2	6	6
Mixed language high	*Particle1-only* (*high frequency*)	1	36	0
	Particle2-only (*high frequency*)	1	0	36
	Alternating	2	6	6

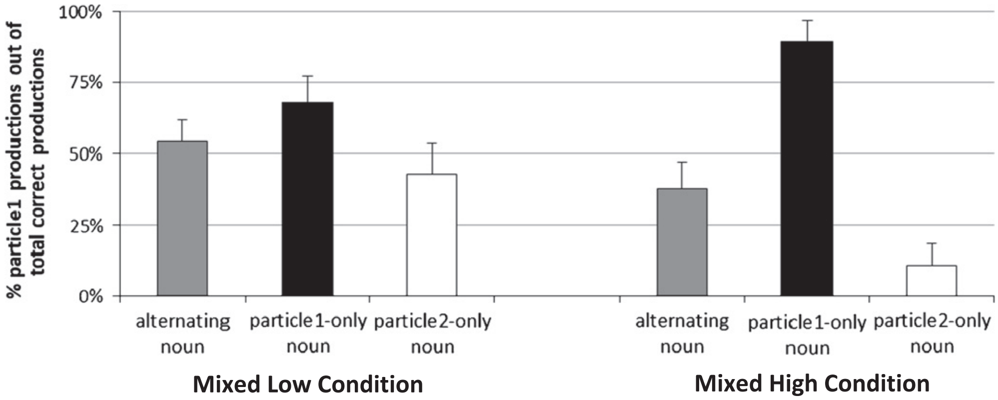

Fig. 4. Experiment 2: Familiar nouns in the MixedLow and MixedHigh conditions. (Note: as all correct sentences have the form (*moop*) *NOUN PARTICLE*, 0% particle1 productions indicates 100% particle2 productions.)

significant learning of the lexical patterns in both conditions, but greater learning in the Lexicalist language, despite the fact that the relevant noun–particle bigrams were matched in frequency across the conditions.

Discussion

The central aim of Experiment 2 was to see whether children could learn that some nouns were restricted to occur with just one of two particles if other nouns alternated between the two, and whether this learning was affected by noun frequency. Two languages were compared which differed only in the frequency of the *particle1-only* and *particle2-only* nouns (and the relevant noun–particle bigrams). The languages were equally well learned in terms of the overall production of (*moop*) *NOUN PARTICLE* sentences, and there was evidence of learning of the restrictions on particle usage in both languages. However frequency did have an effect: there was less generalization with *particle1-only* and *particle2-only* nouns in the language where they were more frequent. This result is in line with the Entrenchment (Braine & Brooks, 1995) and related Statistical Pre-emption (Goldberg, 2005) hypotheses, which suggest that the usage of a lexical item with a particular linguistic structure becomes increasingly unlikely as it is repeatedly encountered in an alternative structure. Here, the usage of a *particle1-only* noun with particle2 is less likely the more that noun has been encountered with particle2 (and vice versa).

In addition to addressing the role of noun frequency, I also compared learning of *particle1-only* and *particle2-only* nouns in the MixedLow condition (where some nouns alternate) with learning of the same nouns in the Lexicalist condition from Experiment 1 (where no nouns alternate). The relevant noun–particle bigrams are matched in frequency, yet learning was significantly stronger in the Lexicalist condition. One possibility is that, as with the *minimal-exposure* nouns in Experiment 1, children are showing their sensitivity to higher-level relationships between nouns and particles – i.e. they are more likely to learn that particular nouns are associated with particular particles when there is greater evidence across the whole language that particle usage is lexically restricted. However, caution must be taken in drawing this conclusion since there is another difference between the two languages: in the MixedLow language the two particles were of equal frequency, whereas in the Lexicalist language particle1 is more frequent than particle2. It may be that this factor also affects learning in some way – for example it might turn out to be easier to learn lexical restrictions in a skewed language. Note, however, that this could *not* explain the results with minimal-exposure nouns in Experiment 1, since the frequency of the particles was matched across the Lexicalist and Generalist input. Moreover, regardless of whether the difference in learning was due to skewed particle frequency, higher level lexical consistency, or some combination of the two, the fact that the relevant bigrams were frequency matched yet differently learned again indicates that children are sensitive to distributional patterns above the level of individual nouns. To this extent, the result corroborates the general findings of Experiment 1.

General discussion

The current study used an artificial language learning paradigm to explore the balance between lexically-based learning and generalization in child language learning. Across two experiments, four groups of 6 year olds were each exposed to one of four semi-artificial languages in which nouns were obligatorily followed by one of two post-nominal particles, but the relationship between specific nouns and particles was manipulated across the languages. In a production test, we looked at which of the two particles children chose to use with different noun types in the different language conditions. Overall, children proved able to produce new noun particle combinations (generalization) but also showed evidence of lexically-based learning – i.e. preferring to produce a noun with the particle with which it had been encountered in the input. Critically, the tendency to replicate lexically-based patterns versus generalize depended on the statistical structure of the input. In Experiment 1 we saw an influence of two "higher-level" statistics: the relative frequency of the particles across the nouns of the input, and extent to which particle usage was lexically conditioned. In Experiment 2 we saw that generalization could also be affected by the frequency of the noun/noun–particle bigram in question. These results demonstrate that, though children are sensitive to item level frequency, aiding the learning of languages where different items behave differently, they are also sensitive to statistical information operating above the lexical level, i.e. above the level of particular nouns or noun–particle bigrams. This corroborates the results with adult learners in Wonnacott et al. (2008), and suggests that this type of learning has relevance for theories of language acquisition.

Wonnacott et al., 2008 suggested that this pattern of learning is consistent with an account where learners (a) track the usage of linguistic structures (here particles) at multiple levels i.e. both for specific nouns and across nouns, and (b) make inferences about the reliability of these different distributions as indicators of future usage. In such an account "lexical conservatism" occurs when the sample of observations for a particular lexical item is deemed a reliable indicator of its continued behavior. This may be formalized in a model which represents degree of belief in a "hypothesis" about particle usage (e.g. in the continued reliability of a distribution) as a probability and uses the mathematics of probability theory to conduct Bayesian inference.

The effect of frequency (seen in Experiment 2) falls naturally out of Bayesian inference since there is a trade-off between the prior bias for a hypothesis and the amount of data that hypothesis captures. The larger the sample of a specific nouns behavior, the more data is captured by a hypothesis based on its past behavior. Note that this is a common property of any model which construes linguistic generalization as rational inference, and thus various Bayesian models capture the effect of linguistic frequency in constraining generalization (see Dowman, 2000 and Chater & Vitanyi, 2007 for related models). Such models may also be considered as formal instantiations of the Entrenchment/Statistical Pre-emption hypotheses.

The results of Experiment 1 suggest that learning can also lead to changes in higher level expectations, so that the probability that particle usage is dependent on the specific noun is deemed greater after exposure to the Lexicalist input (where particle usage was perfectly conditioned on the noun) than after exposure to the Generalist input (where particle usage and noun identity were unrelated). Thus the same noun-specific sample can be treated differently in each case (the trade-off is different). This type of higher-level learning is captured by a specific instance of a Bayesian model – the HBM (hierarchical Bayesian model) which was independently developed by Kemp et al. (2007) and applied to the Wonnacott et al. data by Perfors et al. (2010). Although the precise details of the model are beyond the scope of this paper, a critical feature is that it learns the value of two higher level parameters – "alpha" and "beta" – which constitute expectations about structure/particle usage. The value of "beta" is simply the distribution of the particles across the nouns in the data set (e.g. 0.75 particle1 usage). The value of "alpha" instantiates an "over-hypothesis" about the uniformity of particle usage for particular nouns. If nouns tend to occur consistently with one particle, the value of alpha will be high and lexical distributions will be deemed highly reliable (more lexical conservatism). If particle usage is unrelated to noun identity, the value will be low and lexical distributions will be deemed unreliable (more generalization). In this latter case, the general distribution of particles (beta) may be a more reliable indicator of particle usage. This model (correctly) predicts that children in the Lexicalist condition should tend to produce minimal-exposure nouns with the particles with which they had occurred, while children in the Generalist condition should be more influenced by the language-wide distribution of the particles – i.e. produce both, but particle1 more frequently. It also explains the treatment of entirely-novel nouns: since there is *no* sample for these specific nouns, the language distribution of particles (beta) should influence productions in both conditions, and it did (i.e. children showed greater particle1 than particle2 usage in each case). Learning about higher-level variability (alpha) is also seen in the different degrees of *regularization* with novel nouns across the two conditions: children were more likely to use one particle per novel noun (either the same particle or a different one for each noun) in the Lexicalist condition. This behavior again indicates that children have formed higher level expectations about the relationships between nouns and particles in their input.

Although Bayesian inference in general, and the HMB in particular, provides a neat account of the current data, it is important to consider whether alternative models could capture the same pattern of results. First, it is clear that the results cannot be explained by straightforward associative learning of word co-occurrences, since this would not predict any generalization to new particles. However more powerful models which are capable of generalization may be built from associative learning principles, for example the class of connectionist learning models, which have played an important role in the psycholinguistic literature (e.g. Elman, 1990; Rumelhart & McClelland, 1986; Plunkett & Marchman, 1991). The relative merits of the connectionist and rational inference approaches to cognition is the topic of much ongoing debate (e.g. see Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010 and McClelland et al., 2010). Most relevant here is the question of level of representation. The HBM provides a solution at Marr's computation level: it describes the problem that needs to be solved, and shows us how that problem can be solved if the learner is able to make particular types abstract inferences.

In contrast a connectionist model may provide an account of learning in terms of lower level learning mechanisms. For example, the HBM model assigns values to "alpha" and "beta" by searching the space of all possible pairs of values and determining the candidate which maximize posterior probability – there is no claim our human learners have conducted this type of exhaustive search. In contrast, connectionist models may be trained using methods which are more psychologically plausible. For example, they might learn by predicting upcoming words in each of the exposure sentences (Elman, 1990), allowing them to incrementally update their weights on a trial by trial basis in response to whether their prediction was correct. Like the HBM, connectionist models are able to balance statistical patterns at multiple levels. If a connectionist model were trained on the current languages, it seems likely that it would learn both the associations between particular nouns and particles, and the relative frequencies of each particle across all nouns. Such a model would also be sensitive to sample size (item frequency), and thus capture Entrenchment/Statistical Preemption, since connectionist learning implicitly incorporates a trade-off between the degree of fit with the data and "prior" probability of generalization. (However the latter arises from choices about network architecture, number of training epochs and other network choices, rather than being specified as a prior.) It is currently unclear whether a connectionist model could learn whether particle usage is lexically consistent at a higher level – i.e. the type of learning which accounts for the differing treatment of minimal-exposure nouns across conditions in the Bayesian approach (captured in "alpha" in the HBM). Still, it remains possible that a connectionist, or other more "bottom up" statistical learning model, could capture the data. If so, would its learning approximate the higher level algorithm described by the HBM? This is an interesting challenge for future research. For now, the question of how the relevant computations could be implemented in lower level psychological mechanisms remains open. Nevertheless, the data suggest that the ability to generalize yet avoid inappropriate over-generalization rests on an ability to evaluate the past relevance of lexically-based distributions, i.e. based on sample size and the importance of such information across the language. Note that this provides a solution to Baker's Paradox which does not *rely on* semantic or phonological cues, contrary to the predictions of certain theories (Pinker, 1989) (this does not suggest that these cues do not play a role in natural language learning. I return to this point below).

One potential criticism of the current data is that it depends on production probabilities as opposed to judgments of grammaticality. If resolving Baker's Paradox requires learners to demonstrate their understanding that a form is impermissible in an absolute sense, arguably, this can only be done in a judgment task. Such a test was not included in the current work following pilot work which revealed that in this paradigm many children had difficulty understanding the task (even "child friendly" versions as in Ambridge et al., 2007). However Wonnacott et al. (2008) did include a judgment task as well as a production task, and found that the results of the two were highly correlated. Importantly, like production data, judgment data were also graded, reflecting statistical factors such as frequency and higher-level variability in much the same manner as production probabilities. Wonnacott et al. argued that graded judgment data are in line with the fact that grammaticality judgments in natural languages may also be graded, with more over-generalization with low frequency items, even for adult native speakers (see e.g. Theakston, 2004). The greater overgeneralization in the experimental data was attributed to learners' (far) lesser experience with the artificial language than with a natural language. Similarly, in the current data, there is a good deal of over-generalization, both with minimal-exposure and with the *particle1-only* and *particle2-only* nouns – the point is that the over-generalization is modulated by the statistical properties of the input. The hypothesis is that there would be increasingly less overgeneralization with *particle1-only* and *particle2-only* nouns with increased exposure (even given a language with lots of alternation - although the process would be slower in that case) and that eventually the alterative form would become extremely rare or absent in production, and would also have an increasingly high likelihood of being judged as unacceptable if an appropriate test was included. Another way of thinking about this is that the "Paradox" results from the requirement that learners must *absolutely know* that some non-occurring form is actually ungrammatical, when, logically, it could merely be accidentally absent from the sample of language heard thus far. The current perspective is that whilst it is *logically* true that learners can never be absolutely sure that non-occurring forms may not occur in the future, they can become *extremely certain* that those forms are systematically absent, based on the evidence available in the input. Nevertheless, the questions of how grammaticality judgments and production probabilities are related is important, and future work will continue to explore methods of obtaining such data from children with artificial language stimuli. Ultimately, an important question will be how much evidence is necessary to obtain the extremely consistent judgment data seen with native speakers of natural languages.

A secondary question explored in Experiment 1 was whether children would show evidence of regularization, i.e. creating a system of particle usage where none was evident in the input. One place that we saw this very clearly was with entirely-novel-nouns in the Lexicalist condition. However, as discussed above, children in this condition do have experience of lexicalized patterns of particle usage, albeit with other nouns. In fact the only place where children have the opportunity to show regularization on the basis of *no* evidence is with familiar nouns in the Generalist condition (recall that entirely-novel nouns are tested after the exposure with minimal-exposure nouns, which also provide some lexically consistent input). At first glance the pattern of particle production with these nouns appears in line with probability matching rather than regularization, a phenomenon previously only reported with adults. That is, children use both particles, and with probabilities matching those of the input. However, considering the behavior of individual children, we saw that 33% of children (7/21) produced responses that were regularized in some way. Taking into account the number of productions made by individual children, this was calculated to be significantly more than chance probability matching would predict. Nevertheless, the extent of regularization seems to be less than in previous experiments. For example, given similar language input, Hudson Kam and Newport (2005) found that around 70% of children created some systematic pattern of particle usage, and Austin et al. (2006) found almost 100% of productions used the more frequent particle. One important way in which the current experiments differ from these earlier studies is in the use of a *semi*-artificial language, with nouns borrowed from English. A recent study by Hudson Kam and Chang (2009) suggests that ease of lexical access may increase the tendency to probability match rather than regularize. In that study, adult learners were exposed to an artificial language which they had previously been shown to regularize (Hudson Kam & Newport, 2009) but with tasks which aided lexical retrieval (relevant vocabulary was provided in various ways during the test). They found that with these modified tasks participants showed more probability matching and less regularization than with the standard task. This suggests the possibility that the use of familiar English nouns lead to less regularization than in previous experiments. However there are various other differences between the languages used in the current studies and previous studies, and the different studies use slightly different criteria to establish regularization. Ongoing research is exploring which factors are critical, and also compares children and adults learning the same languages.

Although the central finding of this work is that children's learning is very similar to that of the adult learners in Wonnacott et al. (2008), it would also be interesting if there were differences in the learning of child and adult learners, given the evidence for maturational differences in natural language learning (Johnson & Newport, 1989; Newport, 1990). Unfortunately, differences in the learning paradigms in this study and Wonnacott et al. (2008) (different linguistic structures, numbers of lexical items, etc.) make direct comparison impossible. For example, we have seen that generalization in child and adult learner is affected by the same sorts of statistical considerations, but we cannot say from the current data whether children are any more or less likely to generalize. This question is of some interest given claims in the literature that, at least very young children, tend to be conservative in the early stages of learning (Tomasello, 2000). Other learning studies have found that, though their generalization continues to be evidence-based, children may actually be more conservative than adults given precisely the same input (Boyd & Goldberg, in press; Wonnacott, Boyd, Thomson & Goldberg, 2010). However the language structures used in these studies, and the type of generalization explored, were rather different to the current work, and future work will compare the extent of generalization in adults and children using the current paradigm.

Finally, these results demonstrate that, provided they reliably evidenced in the input, lexically-based patterns of usage are learned even in the absence of any other cues. However in natural language acquisition, syntactic distribution is generally correlated with other lexical properties. Indeed, even where the usage of a form appears to be arbitrary and lexically specified, it often proves possible to identify broader generalizations. For example, subtle functional motivations – semantic or pragmatic – may make the usage of a structure more natural with some lexical items (e.g., Ambridge, Pine, Rowland, & Young 2008; Grimshaw, 1990; Pinker, 1989; Goldberg, 1995); subgroups of lexical items which do/do not occur with a structure may also share some phonological properties (Gropen et al., 1989). Such cues were deliberately excluded from the current languages (and those of Wonnacott et al., 2008) in order to isolatedistributional learning at and above the lexical level. Note that most researchers have concluded that at least some arbitrary lexical specification is necessary to describe natural languages (Braine & Brooks, 1995; Lakoff, 1970; Goldberg, 1995; Boyd, in press); and the current results help to explain the persistence of such idiosyncrasy across generations. On the other hand, it is also clear that adults and older children are sensitive to semantic and phonological regularities, and a full model of learning must explain the interaction of these different sources of information. To that end, future work will explore the learning of new artificial languages which are like those in the current study but where the usage of the two particles is partially correlated with the phonological and semantic properties of the nouns. One possibility is that when the learning paradigm is augmented in this manner, this new learning situation may reveal critical age differences. For example, learners of different ages may differ in how they attend to different sources of information at different stages of learning. Ultimately, the critical question will be whether and how such differences could help to explain the greater success of children at acquiring natural languages.

Acknowledgments

This research was supported by a British Academy Postdoctoral Research Fellowship and an ESRC grant (Award Reference: RES-062-23-0983) awarded to the author. Thank you to Kate Nation, Amy Perfors, Adele Goldberg, Michael Tanenhaus, Kenny Smith, Meghan Clayards and Hei Won Anh for their very helpful comments on various parts of the paper, and to Jennifer Thomson, who collected all the data reported in this paper, and also helped with stimuli creation. Finally thank you to all teachers, parents and children who made the research possible.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jml.2011.03.001.

References

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument-structure overgeneralisation errors. *Cognition, 106*, 87–129.

Austin, A. C., Newport, E. L., & Wonnacott, E. (2006, November). Predictable versus unpredictable variation: Regularization in adult and child learners. *Paper presented at the Boston University Conference on Child Language Development*, Boston, MA.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*, 533–581.

Boyd, J. K. & Goldberg, A. E. (in press). Young children's failure to generalize when exposed to the same input as older learners. *Journal of Child Language*.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review, 70*, 323–348.

Braine, M. D. S. (1971). On two types of models of the internalization of grammars. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 153–186). New York: Academic Press.

Braine, M. D. S., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names of things: Young children's acquisition of verbs* (pp. 353–376). Hillsdale, New Jersey: Lawrence Erlbaum.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76–95.

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*, 720–781.

Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development, 70*, 1325–1337.

Chater, N., & Vitanyi, P. (2007). 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology, 51*, 135–163.

Dowman, M. (2000). Addressing the learnability of verb subcategorizations with Bayesian inference. In *Proceedings of the 22nd annual conference of the cognitive science society*, Mahwah. New Jersey, USA: Lawrence Erlbaum Associates.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*, 179–211.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, B67–B74.

Goldberg, A. E (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2005). *Constructions at work*. Oxford: Oxford University Press.

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

Grimshaw, J. (1990). *Argument structure*. Cambridge, Mass: MIT Press.

Griffiths, T. L., Chater, N. C., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends Cognitive Science, 14*, 357–364.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language, 65*, 203–255.

Hudson Kam, C. L., & Chang, A. (2009). Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 815–821.

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*, 151–195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30–66.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23*, 459–472.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307–321.

Lakoff, G. (1970). *Irregularity in syntax*. New York, Holt: Rinehart & Winston.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg., M. S., et al. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends Cognitive Science, 14*, 348–356.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601–618). Hillsdale, NJ: Erlbaum.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Perfors, A., Tenenbaum, J., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language, 37*, 607–642.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge: MIT Press.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron – Implications for child language acquisition. *Cognition, 38*(1), 43–102.

Rumelhart, D., & McClelland, J. (1986). On learning the past tense of English verbs. In D. Rumelhart & J. McClelland (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, Foundations). Cambridge, MA: MIT Press.

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

Saffran, J., Newport, E., & Aslin, R. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance limitations on grammaticality judgment tasks. *Cognitive Development, 19*, 15–34.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning Memory and Cognition, 19*, 528–553.

Wonnacott, E., Boyd. J. K., Thomson, J. & Goldberg, A. E. (2010). The acquisition of novel constructional patterns by five-year-olds and adults. In *Interactive presentation at the Boston university conference on language development*.

Wonnacott, E., & Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M.R. Clark-Cotton, & S. Ha (Eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development*. Boston: Cascadilla Press.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165–209.