View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Warwick Research Archives Portal Repository

Original citation:
Wonnacott, Elizabeth, Newport, Elissa L. and Tanenhaus, Michael K. (2008) Acquiring and processing verb argument structure : distributional learning in a miniature language. Cognitive Psychology, Vol.56 (No.3). pp. 165-209.
Permanent WRAP url:
http://wrap.warwick.ac.uk/51388
Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.
Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
"NOTICE: this is the author's version of a work that was accepted for publication in Cognitive Psychology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Cognitive Psychology, Vol.56 (No.3). pp. 165-209.
http://dx.doi.org/10.1016/j.cogpsych.2007.04.002
A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk

http://wrap.warwick.ac.uk/

Acquiring and Processing Verb Argument Structure:

Distributional Learning in a Miniature Language

Elizabeth Wonnacott[1]

Elissa L. Newport[2]

Michael K. Tanenhaus[2]

University of Oxford, UK[1]

University of Rochester, NY, USA[2]

Address for Correspondence:

Elizabeth Wonnacott

Department of Experimental Psychology

South Parks Road, Oxford, OX1 3UD.

Telephone: +44 (0)1865 271319

Fax: +44 (0)1865 310447.

Abstract

Adult knowledge of a language involves correctly balancing lexically-based and more language-general patterns. For example, verb-argument structures may sometimes readily generalize to new verbs, yet with particular verbs may resist generalization. From the perspective of acquisition, this creates significant learnability problems (Baker 1979), with some researchers claiming a crucial role for verb semantics in the determination of when generalization may and may not occur (Pinker, 1989). Similarly, there has been debate regarding how verb-specific and more generalized constraints interact in sentence processing (Trueswell et al 1993; Mitchell 1987) and on the role of semantics in this process (Hare et al 2003). The current work explores these issues using artificial language learning. In three experiments using languages without semantic cues to verb distribution, we demonstrate that learners can acquire both verb-specific and verb-general patterns, based on distributional information in the linguistic input regarding each of the verbs as well as across the language as a whole. As with natural languages, these factors are shown to affect production, judgments and real-time processing. We demonstrate that learners apply a *rational* procedure in determining their usage of these different input-statistics and conclude by suggesting that a Bayesian perspective on statistical learning may be an appropriate framework for capturing our findings.

Key Words

Language Acquisition, Sentence Processing, Verb Argument Structures, Eye-tracking, Artificial Language Learning

Introduction

Adult language incorporates both regular, abstract operations and patterns that are idiosyncratic or specific to particular lexical items. The complex interplay between these two types of process is particularly clear in the relationship between verbs and the argument structure constructions in which they may occur. For example, consider the use of the ditransitive structure in English. For many verbs this construction provides an alternative to the use of a prepositional form, as in (1):

(1) Jack gave/brought/threw/ the ball to Henry.

Jack gave/brought/threw Henry the ball.

In addition, the construction may be spontaneously applied to new verbs. For example, Gropen, Pinker, Hollander, Goldberg, and Wilson (1989) demonstrated that children who were taught the new verb *pilk* with the meaning 'transfer by car' would produce such sentences as *"he is pilking him the horse"*. Yet despite this apparent productivity, certain verbs are unexpectedly ungrammatical in the ditransitive, as in *Jack donated/carried/pushed Henry the ball. This illustrates a phenomenon known as *subcategorization*: particular verbs are constrained (or 'subcategorized') as to the set of constructions with which they may occur (Chomsky, 1965). This combination of generalization and lexical restriction turns out to be common across many constructions in different languages (see Pinker, 1989, for a review of the Dative, Causative, Active-Passive and Locative alternations in English), yet poses a puzzle from the perspective of

acquisition. If learners are able to extend verbs to new constructions, how do they learn that some new verb-construction combinations are ungrammatical, rather than simply absent from the particular sample of speech they have heard thus far? This constitutes a significant learning problem, sometimes known as "Baker's Paradox," which has received a great deal of attention in the language acquisition literature (e.g. Baker, 1979; Bowerman, 1988; Braine, 1971; Braine & Brooks, 1995; Brooks & Tomasello, 1999; Pinker, 1989; Theakston, 2004). Although it has been proposed that very young children may avoid the issue by adhering to extreme lexical conservatism (see Tomasello 2000; Fisher 2002a; Tomasello & Abbot-Smith 2002; Gertner, Fisher, & Eisengart, 2006 for review and discussion of the evidence), all researchers agree that, at least from around 3 years of age, children do generalize. When this occurs we also see over-generalization, where children use constructions with verbs for which they are ungrammatical (e.g. **Jay said me no*, Gropen et al. 1989). In this article our focus is on how the retreat from overgeneralization can be achieved i.e. how learners who are able to generalize balance this ability with their knowledge of verb-specific constraints. One obvious source of potential evidence, direct correction from caregivers, has been shown to be very rare in the input to young children (Braine, 1971; Brown & Hanlon, 1970; Newport, Gleitman & Gleitman, 1977). Debate has therefore focused on two alternate sources of evidence: verb semantics and distributional information.

The semantics approach

One approach to a range of difficult problems in syntax acquisition has been to look for semantic or perceptual characteristics that correlate with the syntactic

distributions and to propose that these play a significant role in acquisition (Grimshaw, 1981; Morgan & Demuth, 1996; Morgan & Newport, 1981; Pinker, 1984, 1989). This has sometimes been referred to as the 'bootstrapping' approach, implying that the semantic or perceptual properties are inherently more accessible than the distributional patterns themselves, and therefore might serve as a vehicle by which the distribution can be indirectly acquired. For the current problem, semantic factors would seem to be particularly promising, since there is a strong correlation between a verb's meaning and the set of structures in which it may occur (Fisher, Gleitman, & Gleitman, 1991; Gleitman, 1990). Most fundamentally, the use of a particular argument structure imposes constraints on the number of arguments that must be associated with the verb. For example, causal events generally require structures with two NP slots, and transfer events require structures with three NP slots. Furthermore, the use of a particular structure may also have more subtle semantic connotations[1], for example the use of the ditransitive implies not only transfer but also transfer of possession (Green 1974; Goldberg, 1995; Jackendoff, 1972; Levin, 1993; Pinker, 1984, 1989).

Young children show an awareness of these correlations between meaning and structure in their usage of new verbs (Gropen, Pinker, Hollander & Goldberg, 1991a,b; Gropen, Pinker, Holander, Goldberg & Wilson, 1989), and many researchers agree that these semantic properties play some constraining role in the acquisition of verb distribution (e.g. Bowerman 1988; Braine & Brooks, 1995; Pinker, 1984,1989). Pinker (1989) takes the bootstrapping approach further, proposing that children acquire a complex system of semantically and morpho-phonologically defined 'narrow' verb classes which determine syntactic subcategorization. (For example, *carry, push* and *lift*

belong to a class of verbs which do not occur in the di-transitive construction and which share the meaning *'continuous parting of force in some manner causing accompanied motion'*.) However, several researchers have pointed out that some of the postulated class criteria are inconsistent, so that they do not capture the full pattern of verb-structure co-occurrences (Bowerman 1988; Braine & Brooks, 1995; Goldberg 1995, chapter 5). The general conclusion is that, although there are strong correlations between the two, verb distribution cannot be reduced to verb semantics (or a combination of semantic and perceptual cues). Moreover, experimental evidence suggests that, even when the generalizations captured by Pinker's classes are real and productive in the adult grammar, young children may only acquire this semantic knowledge relatively late in development (Ambridge, Pine, Rowland and Young, in press; Brooks & Tomasello, 1999).

More generally, a large body of research highlights the complexities involved in acquiring verb semantics at any developmental stage – specifically, the difficulty in extracting verb meanings from purely environmental contingencies (Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Gillette, Gleitman, Gleitman, & Lederer, 1999; Snedeker, 2000). In fact, Gleitman and colleagues have argued that the problem is so hard as to be intractable for many verbs. Instead, they propose that the acquisition of verb semantics itself relies on a process of distributional learning, i.e. that learners make use of information about the linguistic structures in which a verb has occurred to make inferences about its meaning – a process sometimes called 'syntactic bootstrapping' (Gleitman, 1990). One piece of evidence for this process is that young children are able to use the syntactic frame in which a new verb occurs to make inferences about its meaning. (Naigles, 1990; Fisher, 1996, 2002b).

This line of research clearly challenges any account in which prior learning of verb-semantics provides the crucial 'bootstrap' for acquiring verb-syntax. In addition, the syntactic bootstrapping literature provides a further motivation for the current work. Gleitman and colleagues have claimed that *"the set of frames associated with single verbs provides convergent evidence as to their full expressive range"* (Gleitman et al., 2005; see also Fisher et al., 1991). Thus an important component of that theory is that learners are able to acquire verb-structure co-occurrences independent of verb meaning. The current work investigates this learning process.

The statistical approach

An alternative to using semantic correlates to acquire verb distribution is to extract that information directly from the set of verb-structure combinations occurring in the input. Such a theory was first proposed by Braine (1971) and has recently gained in popularity (e.g. Braine & Brooks, 1995; Goldberg, 1995; 2007; Theakston, 2004; Tomasello 2000). Two potential learning processes have been discussed: *pre-emption* and the *entrenchment*.

Pre-emption refers to the evidence provided by encountering a verb in construction A when construction B would have provided the same communicative function. The hypothesis is that the pre-emption of B provides evidence that it is ungrammatical with that verb. (Related hypotheses have been propounded for many aspects of language acquisition, e.g. Markman (1989)'s principle of mutual exclusivity for word learning, Pinker (1984)'s uniqueness principle for morphology). Goldberg (1995) suggests that this process is aided by the fact that the form which is actually

encountered will often be less felicitous than the form it preempts (given that no two constructions are entirely synonymous, Givon, 1985). For example, since the periphrastic causative is not generally associated with direct causation, the sentence *'He made the rabbit disappear'* should be less felicitous than *'*He disappeared the rabbit'*. Thus encountering the former sentence in place of the latter could provide the child with evidence that *disappear* is ungrammatical in the transitive. There is evidence for the use of this information, but so far only in older children (4.5 year olds, Brooks & Tomasello, 1999; 6 year olds, Brooks & Zizak, 2002).

In contrast, *entrenchment* has been construed as blind to the semantic or pragmatic properties of verbs or constructions. The notion is that encountering a verb frequently in the input 'entrenches' its use with the particular constructions with which it has occurred, making it less likely to be generalized for use with a new construction (Braine & Brooks, 1995). A number of studies have found evidence for entrenchment in child language. Tomasello, Dodson and Lewis (1999) found that three year olds were more likely to produce over-generalizations with low frequency English verbs (*he arrived me to school) than high frequency equivalents (*he came me to school). Similarly, Theakston (2004) and Ambridge et al. (in press) found that five year olds were likely to give such over-generalizations higher grammaticality ratings when they involved low frequency verbs. Mathews, Lieven, Theakston and Tomasello (2005) explored the ability of 2;6 year olds to use known English verbs with a new construction introduced during the experiment (SOV word order), finding that they were more likely to produce such generalizations with low frequency than high frequency verbs.

Although there is some question as to whether entrenchment and pre-emption are really distinct (see Goldberg, 2005, for the argument that frequency effects reflect the verb's more frequent occurrence in specific pre-empting constructions), these phenomena provide evidence that learners are sensitive to the frequencies of various combinations of verbs and structures occurring in the input and that they use this information to make inferences as to the status of 'missing' verb-structure pairs. This account concurs with a growing body of research demonstrating that learners come equipped with powerful *statistical learning* mechanisms. For example, Saffran, Newport and Aslin (1996; Saffran, Aslin & Newport, 1996) showed that adults and young infants could track the frequencies and conditionalized probabilities of syllable co-occurrence patterns in a speech stream and apply these statistics to the process of word segmentation; Mintz (2002) showed that adult learners could abstract syntactic categories on the basis of word co-occurrences. The current work explores whether statistical learning can be extended to the problem of verb argument structure acquisition. We also explore whether this same learning process can account for related phenomena in real-time sentence processing.

Verb argument structures in sentence processing

Verb subcategorization has also featured in a parallel literature on on-line sentence processing. Here the focus has largely been on statistical rather than absolute constraints (sometimes known as a verb's subcategorization 'profile'). For example, many verbs in English may be followed by a choice of complement structures but are constrained as to how readily they occur with these structures (for instance, *find* may occur with either a direct object or a sentential complement, as in *Arthur found Trillian*

and *Arthur found Trillian was in the car*, but is more likely to occur with a direct object). In this literature too, debate has focused on role of this verb-specific knowledge versus more generalized patterns. One influential approach has proposed that real-time comprehension is primarily influenced by biases which operate above the level of individual verbs (Frazier, 1987; Frazier & Fodor, 1978; Frazier & Rayner, 1982). According to this theory, these biases arise from an inherent preference for syntactic simplicity (embodied in various parsing principles such as *Minimal Attachment*), which influences the structures assigned during real time parsing, irrespective of the particular lexical items involved. Although early experimental work appeared to support this hypothesis (Clifton, Frazier & Randall, 1983, Ferreira & Henderson, 1990; Mitchell, 1987), a large body of work now indicates that verb-specific biases have a strong and immediate influence in real-time processing (Garnsey, Pearlmutter, Meyers, & Lotocky, 1997; Snedeker & Trueswell, 2004; Trueswell & Kim, 1998; Trueswell, Tanenhaus, & Kello, 1993; but cf. Kennison 2001). For instance, Trueswell, Tanenhaus, & Kello (1993) investigated whether readers were sensitive to the likelihood of particular English verbs being followed by a either a direct object or a sentential complement. Participants' eye-movements were monitored as they read pairs of sentences such as: *'The chef found/claimed (that) the recipe would require using fresh basil'.* Reading times suggested a bias to interpret the post-verbal NP (*the recipe*) as a direct object with direct object biased verbs like *find*, but not with sentential complement biased verbs like *claim*, indicating that subcategorization information was accessed and used to determine upcoming structure as soon as the verb was processed. Similar lexical effects have also

been found in spoken language comprehension, with both adults and children (Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill & Logrip, 1999).

Although the influence of lexically-based, verb-specific biases is now well established, there is evidence that more abstract, verb-general biases also play a role in processing. In particular, it has been shown that post-verbal nouns are occasionally interpreted as direct objects, even with verbs which have never occurred with that type of complement (Mitchell, 1987; Juliano & Tanenhaus, 1993). This phenomenon suggests a verb-independent structural bias, which has been attributed to the inherent preferences of the parsing architecture (Frazier, 1987; Mitchell, 1987). However, Juliano and Tanenhaus (1993, 1994) argued that these effects occur primarily with low frequency verbs. This effect of frequency, akin to the process of entrenchment discussed in the developmental literature, would again appear to signature a statistical process. Juliano and Tanenhaus suggest that the bias may arise from the general preponderance of that type of complement structure *across* the verbs of the language. This hypothesis has been explored in a number of computational models (Juliano & Tanenhaus, 1994; Tabor, Juliano & Tanenhaus, 1997; Kim, Bangalore & Trueswell, 2002). Critically, these models track both verb-specific statistics (the likelihood of particular verb occurring in a particular structure) and verb-general statistics (the occurrence of different argument structures, across verbs in the language).

Despite the apparent success of these statistical accounts, it is not possible to conclude that even verb-specific biases are actually distributional in nature. A recent series of studies suggests an important role for verb semantics (Hare, McRae, & Elman,

2003, 2004). This work demonstrates that the subcategorization preferences which play a role in online processing are sense contingent. For example, for the verb *find*, its *locate* sense is subcategorized to occur only with a direct object, and processing is sensitive to this information (Hare et al., 2003). One interpretation of these results is that distributional analyses are performed not over particular lexical forms, but over particular senses of those forms. However, these findings at least raise the possibility that structural preferences may be entirely driven by verb-semantics. The strong correlations between verb distribution and verb semantics, which hold in each natural language, make it impossible to determine whether verb biases are a result of the verb's own distributional history or of its membership in some more general semantic classes.

One method of avoiding the confounds inherent in natural language is to explore the learning of artificial languages, in which these factors can be disentangled. One previous study found evidence that learners could acquire and use probabilistic subcategorization patterns which were entirely distributional in nature. Wonnacott and Newport (2005) exposed learners to an artificial language in which all verbs could occur with either of two constructions, but occurred with one construction twice as often as the other. In contrast to natural language input, there were no semantic or structural reasons to prefer the use of any construction with any verb. Participants were then asked to produce their own sentences in the language. The central finding of this study was that the tendency to use each construction with any verb matched the probabilities of the input: participants used the dominant construction twice as often with each verb. In the experiments reported below, we extend this methodology to ask whether learners can acquire both verb-specific and verb-general distributional information, and how the

distributional nature of the input influences learners' usage of these different input statistics.

The current work

We explore the hypothesis that the subcategorization phenomena reported in the acquisition and processing literatures can be accounted for by *statistical learning* processes (cf. Saffran et al., 1996; Mintz, Newport & Bever, 2002; Thompson & Newport, 2007) -- that is, that learners track the occurrences and co-occurrences of verbs and structures in the input, and can use that information in a sophisticated way to make inferences about the underlying language system. Our approach suggests that the problem of learning when to restrict constructions to specific verbs, versus generalize their use, is part of a larger process of balancing verb-specific and verb-general statistical information.

The benefits of using artificial languages as a means of obtaining precise control over the input to learning is now well established (Aslin, Saffran & Newport, 1998; Braine, 1963; Hudson Kam & Newport, 2005; Gerken, 2006; Gomez, 2002; Mintz, 2002; Moeser & Bregman, 1972; Morgan & Newport, 1981; Morgan, Meier & Newport, 1987; Saffran et al., 1996; Wonnacott & Newport, 2005). In addition, there is emerging evidence that artificial languages exhibit many of the same signature results in processing as those obtained with natural language stimuli (e.g., Magnuson, Tanenhaus, Aslin & Dahan, 2003). Here the methodology allows us to create languages in which the relationship between verbs and potential argument structures is entirely distributional in nature (there are no structural reasons to prefer the use of any argument structure and no

semantic or phonological correlates to verb behavior), and then to manipulate verb-structure co-occurrences across different artificial languages, to observe how different distributional patterns affect learning.

In order to explore not only whether learners acquire different distributional relationships, but also how they use that information, all of the experiments involved three different language tests: *grammaticality judgment*, *production*, and *online comprehension*. Although most artificial language experiments have relied on grammaticality judgments (Braine 1990; Gomez, 2002; Moeser & Bregman, 1972; Morgan et al., 1987; Morgan & Newport, 1981; Saffran et al., 1996), some studies have also included tests of production and online comprehension (production: Hudson-Kam & Newport, 2005; Wonnacott & Newport, 2005; online comprehension: Magnuson et al., 2003). Including all three measures in the same experiments will allow us to compare different modes of learning. As in Wonnacott and Newport 2005, the production test was set up to ascertain the tendency to produce each of the possible constructions with each of the verbs in the language. The function of the online comprehension test was to test the tendency to predict each of the constructions' likelihood with different verbs. To that end, this test employed eye-tracking in the Visual World Paradigm (Altmann & Kamide, 1999; Cooper, 1974; Tanenhaus et al., 1995).

Each of the four languages used in these experiments had the same basic vocabulary of nouns and verbs, and involved the same two argument structures (*Verb Agent Patient* and *Verb Patient Agent Particle*). What was manipulated across the

different languages was whether and how often different verbs occurred in each of the two structures, with no semantic or other cues to verb distribution in any language.

In Experiment 1 we investigate whether participants can acquire verb-specific constraints (i.e. learn that certain verbs can occur in only one of two competing constructions), even in the presence of a class of unconstrained 'alternating' verbs that occur in both constructions. We also ask whether, as in natural language learning and processing, these results are modulated by verb frequency. In Experiment 2 we ask whether we can tip the balance between acquiring verb-specific constraints and generalizing by manipulating the distribution of verb types across the language as a whole. We expose learners to a language containing a larger and more varied class of alternating verbs than in Experiment 1, asking whether this makes them more likely to generalize alternating frames to the constrained verbs. In Experiment 3, we expose learners to languages in which verb-specific and verb-general patterns are probabilistic and examine the influence of these statistics on the different language behaviors. We ask what happens when these different statistics are in conflict, and whether this conflict is affected by the overall distributional properties of the language.

Taken together, these experiments will allow us to determine whether distributional learning mechanisms are able to acquire the types of lexical constraints discussed in the acquisition and processing literatures, and also how the distributional details of the input influence the balance between applying lexical patterns and generalizing.

Experiment 1

The aim of Experiment 1 was to investigate whether participants could learn verb-specific construction constraints -- that is, to avoid over-generalizing with verbs that only occurred in one of two constructions when the language also contained a class of verbs that occurred in both constructions. To this end, we exposed learners to a language with three verb classes: a class of *alternating* verbs (occurring equally often in each of the two possible constructions) and two classes of *one-construction* verbs (each occurring only in one of the two constructions). This set-up was intended to resemble the pattern seen with alternating constructions in natural languages. However, there were no semantic or phonological cues to verb class. Thus in this experiment participants could only acquire verb subcategorization restrictions by paying attention to the actual verb-structure co-occurrence statistics. Although all three classes of verbs were equal in size and frequency, we also manipulated individual verb frequency within the one-construction verb classes. We predicted that we would see frequency effects mimicking those seen in natural languages, i.e. that participants would be less likely to produce, accept and predict over-generalizations with high frequency verbs than with low frequency verbs.

Method

Participants

Fourteen adult native English speakers participated in the experiment. Most were undergraduate students at the University of Rochester; the remainder were former students or age mates who attended another college. Participants were run and tested

individually. They were paid daily for their participation and received a bonus upon completion of the entire experiment.

Description of the Language

The language had the following features:

Vocabulary

Nouns: There were five two-syllable nouns, each beginning with a different initial consonant and referring to 5 puppet animals [*flugat* (BEE), *blergen* (LION), *slagum* (LADYBUG), *nagid* (ELEPHANT), *tombat* (GIRAFFE)]. (The vocabulary set is adapted from that used in Hudson Kam & Newport, 2005). In contrast to verbs, referents were fixed across participants. There were no lexical restrictions on nouns; each occurred freely with all verbs and in each of the verbal structures.

Verbs: There were 12 one-syllable verbs [*glim, norg, frag, flern, semz, gund, loom, mer, rov, shen, gofe*] denoting 12 transitive actions [PUSH, STROKE, TICKLE, ROCK, KISS, HUG, RAM, BOP, JUMP-ON, DRAG, SIT-ON]. The assignment of verbs to actions was randomized across participants (as was the assignment of verbs to the verb classes – see *Verb Classes*, below).

Particles: There was one particle word '*ka'.* This had no referent but occurred in one of the two constructions and could be regarded as a linguistic marker of that construction.

Grammar

The language contained two possible sentence constructions:

Verb **Noun1** Noun2	(e.g., *glim tombat blergen*)
Verb Noun2 **Noun1** Particle	(e.g., *glim blergen tombat ka*)

Both of the two constructions corresponded to a simple transitive event structure in which *Noun1* was the agent and *Noun2* the patient of the act denoted by the verb. (Henceforth we refer to the two constructions as the *VAP* (verb agent patient) and *VPA_ka* (verb patient agent 'ka') constructions). For example, if *glim* means HIT, *tombat* means GIRAFFE, and *blergen* means LION, then:

glim tombat blergen	means	THE GIRAFFE HIT THE LION
glim blergen tombat ka	also means	THE GIRAFFE HIT THE LION

These forms were designed to have the following crucial properties:

Synonymy. Although it would be unusual, perhaps impossible, to find two entirely synonymous forms in any natural language (Givon, 1985), our aim in this work was to test whether syntactic constraints can be learned in the absence of semantic or pragmatic cues. Therefore in these experiments VN_1N_2 and VN_2N_1*ka* were synonymous, so there could be no functional reason for preferring either construction for any specific verb.

Temporary ambiguity. Sentences in this language were potentially ambiguous until the final particle. That is, if a learner hears *Verb Noun Noun …,* she will not know if this was a *Verb Agent Patient* or *Verb Patient Agent* sequence until the final *ka* particle is (or isn't) heard. Crucially, however, this ambiguity holds *only if the language learner believes that the verb can occur in both constructions.* Thus, if we are able to access

participants' expectations about the structure they are hearing *before* they hear the disambiguating *ka* particle, we may learn: (a) whether they have acquired verb subcategorization and (b) whether they are able to use this knowledge in real-time sentence comprehension.

For example, imagine a participant is hearing the sentence:

glim flugat blergen ka

RAM BEE LION PARTICLE

as it unfolds. If the learner has acquired the constraint that *glim* can occur only in the *VPA_ ka* construction, and if he or she is able to apply this constraint online, then as soon as the sentence begins and the verb is identified, the *VPA_ ka* construction (and not the *VAP* construction) should be activated. Each noun begins with a distinct consonant. Thus as soon as the first noun is identified, the learner will expect this to be the patient. In the present example, this will be at the point where she has heard *glim f…*. At this point the learner should predict a sentence semantics in which BEE is the patient of the event RAM. On the other hand, if the learner thinks that *glim* can occur in both constructions, then the *VAP* and *VPA_ka* constructions will both be activated. In this case, the participant will not assume that BEE is the patient of RAM action until the final *ka* particle has been heard.

Verb Classes. The 12 verbs were divided into three classes:

- 4 *VAP*-only verbs (verbs that always occurred with the *VAP* structure)
- 4 *VPA_ka*-only verbs (verbs that always occurred with the *VPA_ka* structure)

- 4 alternating verbs (verbs that occurred with each structure 50% of the time)

Henceforth we will refer to *VAP*-only and *VPA_ka*-only verbs as one-construction verbs.

The two one-construction verb classes were then further sub-divided into low frequency (2 verbs) and high frequency (2 verbs). Low frequency verbs occurred in the input set less often than the alternating verbs, and the high frequency verbs occurred more often than the alternating verbs (see *Input Sentence Sample*).

To make sure there was no accidental preferred association between any verbs and their construction classes, a different random assignment was performed for each participant. This ensures that there are no unnoticed regularities between verb class and semantic or phonological features.

Input Sentence Sample. The input set in this experiment had the following distributional properties:

- *VAP*-only, *VPA_ka*-only and alternating verb classes occurred equally often.
- *VAP* and *VPA_ka* constructions occurred equally often.
- Within the verb classes, verbs occurred equally often, except that high frequency verbs occurred three times as often as low frequency verbs and 1.5 times as often as alternating verbs.

Materials

Sentences were built from a set of pre-recorded sound files for each of the individual words. Each of the 12 verbs, five nouns, and the particle were recorded with an

intonation pattern appropriate to each of the possible sentence positions in which it could occur, and the resulting sound files were edited so that words of each type were of a standard duration (verb 470 ms, noun in first position 570 ms, noun in second position 625 ms, particle 470 ms). The words were concatenated into sentences in real-time during the experiment. This methodology ensured that word boundaries occurred at the same point across sentences, and also that two sentences such as *glim tombat blergen* and *glim tombat blergen ka* were identical up until the final *ka* particle. Participants questioned after the experiment seemed unaware that sentences were not produced as whole units. These audio stimuli were played along with video clips which provided the semantics. The clips involved scenes in which pairs of puppets engaged in simple transitive events (e.g. GIRAFFE HIT LION). The experiment ran using ExBuilder, a custom software package. The software was responsible for concatenating sentences from the word sound files and performed the participant-by-participant vocabulary randomization and the assignment of verbs to verb classes.

Procedure

A 5-day procedure was used for this experiment. One procedure was followed on days 1-3 (which were primarily for sentence exposure) and another on days 4 and 5 (which were primarily for testing).

Procedure Days 1 to 3.

Part 1 - Vocabulary Learning: Participants viewed static pictures of each of the five puppet animals while hearing their names (4 exposures per noun), and then took a

short vocabulary test in which they saw each picture and produced the appropriate noun orally. They then received feedback and help from the experimenter. This process was then repeated. Most participants could name every animal correctly by the second vocabulary test on day 1, and all participants could do this on day 2.

Part 2 - Sentence Exposure: Participants viewed a set of 144 transitive scene sentence pairs. The input set had the structure described above, and sentences were presented in random order. Participants were instructed to repeat the sentence out loud to help themselves learn.

Part 3 - Short Tests: These tests were mainly included to give participants some experience of the test procedures and were not scored. In order to assess whether test exposure would dilute the frequency manipulation, half of the participants were not given this part of the experiment on days 1 and 3. However, there were no differences between these groups, so the participants are all included in all analyses. Participants took the three tests in the order described below. Test items were presented in random order in each test.

Production Test: In each test trial, participants saw a scene and heard the verb corresponding to the action in the scene, followed by a pause. Their task was to complete the sentence. This method, which follows the procedure established by Hudson Kam and Newport (2005), allows participants to demonstrate knowledge of sentence structure and verb class without having explicitly memorized each of the twelve verbs. There were 12 test items, one for each verb.

Online Comprehension Test: In each trial participants viewed two video scenes showing the same pair of animals and the same event, but with reversed roles taken by the agent and patient (e.g. video 1: BEE RAM LION, video 2: LION RAM BEE). They then heard a sentence in the language (e.g. *glim flugat blergen ka*) and had to choose which scene depicted the correct meaning of the sentence. There were 12 test items, one for each verb. In order to minimize interference with the input statistics, all of the sentences we used obeyed the verb subcategorization restrictions that appeared in the input set: *VAP*-only verbs occurred with the *VAP* construction, and *VPA_ka*-only verbs occurred with the *VPA_ka* construction. Two of the four alternating verbs occurred with the *VAP* construction, and two occurred with the *VPA_ka* construction.

On days 1-3 we did not record eye movements during the Online Comprehension Test. However, the purpose of these tests was to familiarize participants with the test format that would be used on days 4 and 5. Thus trials included specific procedures for re-calibration and centralizing eye-fixations. The structure of each trial was as follows:

A white cross appeared in the center of the screen, which participants clicked on to start the trial. (When the test was eye-tracked, an automatic drift correction of the calibration was performed at this point). The two videos then appeared side by side on the screen, frozen on the first frame (see figure 1). Video 1 played, followed by Video 2, and the two pictures remained frozen on the screen in final frame. Another white cross appeared between the two pictures, and participants clicked on this cross to play the sentence (this centralized fixations, ensuring that participants were not looking at either picture before the sentence was heard). The sentence was heard. Participants then chose

a picture by using the mouse to click on a red button at the bottom of that picture. They could only do this after the sound file had finished playing. Choosing the picture ended the trial.

FIGURE 1 ABOUT HERE

Grammaticality Judgment Test: Participants viewed a video scene and heard a sentence with either a *VAP* or a *VPA_ka* structure. There were 12 test items, one for each verb. For one-construction verbs, half of these sentences broke the subcategorization restrictions of the input sentences, i.e. two *VAP*-only verbs occurred with *VPA_ka*, and two *VPA_ka*-only verbs occurred with *VAP*. Two of the four alternating verbs occurred with the *VAP* construction and two occurred with the *VPA_ka* construction. Participants were asked to make two responses, one after the other:

- Is this a correct sentence in the language? [Participants click 'Yes' or 'No']
- Give the sentence a score out of 5, where 5 means 'definitely a correct sentence in the language' and 1 means 'definitely not a correct sentence in the language'. [Participants click on a number between 1 and 5].

Measures of both binary judgment and ratings were included on the assumption that they would show differences in sensitivity. However, no such differences were found. Since the two measures yielded an identical pattern of results in all of the experiments, the ratings data are not included in this paper.

Procedure on Days 4 and 5.

Part 1 - Vocabulary Test: Participant saw pictures of each animal and said the appropriate noun. All participants scored 100% on both days.

Part 2 - Intermittent Short Exposure and Longer Tests.

Short Sentence Exposure 1: Participants viewed a subset of 72 sentences from the exposure set used on the previous days. Again they were instructed to repeat the sentence out loud to help themselves learn.

Production Test: This was exactly like the production test given on days 1-3, except that there were 24 test items, 2 per verb, and all the scenes used in this test were entirely novel, i.e. had not occurred in the sentence exposure or the previous production test. (This ensured that participants could not base their productions on specific memorized sentences.) We scored the percentage of productions which correctly applied one of the two possible sentence constructions, and the proportion of these which corresponded to each construction for each verb.

Short Sentence Exposure 2: Participants viewed a subset of 24 sentences from the exposure set. (This was included between tests to refresh participants' knowledge of the language and to minimize test effects on this knowledge).

Online Comprehension Test: This was like the Online Comprehension Test given on days 1-3, except that there were now 24 test items, 2 per verb (sentences with one-construction verbs obeyed the verb's subcategorization restrictions, and each alternating verb occurred once with each construction); all of the sentences used were entirely novel, i.e. had not occurred in sentence exposure or the previous Online Comprehension Test;

and participants were now eye-tracked while listening to the sentences and viewing the two scenes.

Eye-movements were monitored using an Eyelink II head mounted eye-tracker. At the beginning of the Online Comprehension Test, the system was mounted and the eye-tracker calibrated using the standard nine-point Eyelink calibration procedure. As described above, the trials included components which allowed for drift correction and centralized fixations. Eye-movements were monitored from the point at which the sentence began playing until the participant finished the trial by clicking on the red button at the bottom of one of the pictures (only possible after the sentence had finished). The Eyelink software automatically parses the eye track into three categories: saccades, fixations, and blinks. A look to one of the two pictures was defined from saccade onset to the end of the fixation. Because participants tended to look slightly below the picture when they clicked on the red button, the bottom border was extended by 15 pixels to capture more fixations. We recorded the average proportions of looks to each of the two pictures every 4ms, as the sentence unfolded in time.

Short Sentence Exposure 3: Participants viewed another subset of 24 sentences. (This was included between tests to refresh participants' knowledge of the language and minimize tests effects on this knowledge).

Grammaticality Judgment Test: This was exactly like the test on days 1-3 except that there were now 24 test sentences, two per verb; and all of the sentences used in the test were novel. Each verb occurred with each of the two constructions, so that half of the sentences with one-construction verbs broke the subcategorization restrictions. We

recorded whether participants judged each verb to be grammatical in each of the two constructions.

Results

Overall, 98% of productions were found to have word orders conforming to one of the two possible constructions (*verb agent patient* or *verb patient agent particle*), showing that participants had indeed mastered the basic language structures[2]. The small minority of incorrect sentences had one of two other word orders: *verb agent patient particle* or *verb patient agent*. These were excluded from the analyses that follow.

Figure 2 shows, for each verb class, the proportion of correct productions which used the *VAP* construction (since incorrect productions are excluded, all remaining productions used the *VPA_ka* construction). In these and all subsequent experiments, there were no reliable effects of *Practice Test Exposure* (whether participants had taken the extra practice tests on days 1 and 2) or *Day of Testing* (whether the test was taken on day 4 or day 5), and these factors did not interact with the effects of either *Verb Class* or *Verb Frequency*. These factors are therefore collapsed.

FIGURE 2 ABOUT HERE

Overall, 58.8% of productions were produced with the *VAP* construction, a marginally significant bias towards that construction (t=1.8, p<0.1, df=13). However, in spite of this bias, participants demonstrated clear knowledge of the subcategorization restrictions on the one-construction verbs. The majority of productions with *VAP*-only verbs used the *VAP* construction, and the majority of productions with *VPA_ka*-only

verbs used the *VPA_ka* construction. Alternating verbs used each of the two constructions approximately equally often (the slight bias towards the *VAP* construction was not significant within this class, t=0.75, df=13, p>0.05). Overall, the main effect of verb class was significant (F(2,24)=16.31, p<0.01), and planned comparisons showed significant differences between all three classes (*VAP*-only versus *VPA_ka*-only: t=4.9, $p<0.01$, df=13, *VAP*-only versus alternating: t=3.76, $p<0.01$, df=13, and *VPA_ka*-only versus alternating: t=2.49, $p<0.05$, df=13).

As predicted, this learning of verb restrictions was modulated by verb frequency: participants used the correct construction more often with high frequency verbs than with low frequency verbs, F(1,12)=5.12, p<0.05. This effect mirrors so-called *entrenchment* effects in natural languages: participants are less likely to use verbs in new structures (i.e., to produce overgeneralizations) when those verbs have occurred more frequently in the input.

Grammaticality Judgment Test

Figure 3 shows, for each of the different verb types, the difference between the grammaticality judgments given to sentences using the *VAP* construction and those using the *VPA_ka* construction. (These scores are calculated by subtracting the judgment scores for each construction type. These raw data are shown in Table 1.) Note that the scores are with respect to the *VAP* construction, so that a score of 100% would mean always accepting *VAP* sentences and rejecting *VPA_ka* construction sentences, -100% would mean always accepting *VPA_ka* sentences and rejecting *VAP* sentences, and 0% would mean judging the two constructions to be equally grammatical.

FIGURE 3 ABOUT HERE

TABLE 1 ABOUT HERE

The difference scores show a significant main effect of verb class ($F(2,24)=16.27$, $p<0.01$). Participants were more likely to accept the *VAP* construction with *VAP*-only verbs and more likely to accept the *VPA_ka* construction with *VPA_ka*-only verbs, whereas with alternating verbs they accepted the two structures approximately equally often. Planned comparisons showed significant differences between all three classes (*VAP*-only versus *VPA_ka*-only: $t=4.39$, $p<0.01$, $df=13$, *VAP*-only versus alternating: $t=3.84$, $p<0.01$, $df=13$, and *VPA_ka*-only versus alternating: $t=3.3$, $p<0.01$, $df=13$).

For the two one-construction classes, we also observed effects of verb frequency ($F(1,12)=6.81$, $p<0.05$). Turning from the difference scores to the scores for the different construction types (see Table 1), we can see that this is largely due to the greater tendency to reject *incorrect* overgeneralizations: for *VAP*-only verbs: 67.9% of low frequency and 39.4% of high frequency verbs were accepted in sentences with the *VPA_ka* construction; for *VAP*-only verbs: 46.4% of low frequency and 55.4% of high frequency verbs were accepted in sentences with the *VAP* construction. Again, this mirrors the entrenchment effects which have been described in the developmental literature.

In contrast to the results of the other tests, we see no bias towards the *VAP* construction in these data (in fact there were 2% more *VPA_ka* productions overall, $t=-0.76$, $df=13$, $p>0.05$).

Online comprehension test

Overall participants chose the correct picture on 95.7% of trials, showing again that they had mastered the two possible word orders in the language. However, the purpose of this test was to ascertain whether participants were able to use their knowledge of verb subcategorization to form expectations in real time processing. That is, we asked whether looking preference would reveal different points of disambiguation for the one-construction and alternating verbs. Figure 4 shows the proportion of looks to each of the two pictures as participants heard sentences using each of the two constructions.

FIGURE 4 ABOUT HERE

The results are particularly clear when we compare looking patterns as participants hear *VPA_ka* sentences that begin with *VPA_ka*-only versus *VPA_ka* sentences that begin with alternating verbs. When the sentence begins with a *VPA_ka*-only verb, participants begin to look more often at the picture corresponding to the *VPA_ka* sentence at around 916ms, i.e. during the first noun. In contrast, when the sentence begins with a alternating verb, looks to the *VPA_ka* picture do not begin to rise until around 1900, i.e. *after* the participant has heard the disambiguating *ka* particle (before that time, in the critical region there is a bias towards the *VAP* construction). This demonstrates that participants show verb-specific biases online, and that, as in natural languages, this information is used at the earliest possible moment in processing.

For *VAP* sentences the comparison is less clear. In both cases, there are more looks to the *VAP* picture in the critical region. In the case of the alternating verb, this again demonstrates the inherent bias towards the *VAP* construction. However, in both cases there are, by chance, more looks to the *VPA_ka* picture in the initial period, causing looks to the *VAP* picture to rise even before the first noun has been heard (at which point there is no reason to identify either picture with either of the two constructions). Thus it is not possible in either case to identify the true point of disambiguation.

In order to factor out the effects of previous looking behavior, and to allow easy comparison across the different verb types, we also counted the total proportion of looks to each picture which *began* during the critical region from the beginning of the first noun (when they first potentially have enough information to identify the sentence structure), up to the beginning of the *ka* particle (the final disambiguating information). Allowing the standard 200ms for programming an eye-movement (Hallet, 1986), this period was taken to be the time period 669-1850ms. These data are shown in Figure 5.

FIGURE 5 ABOUT HERE

There is a clear influence of verb type on looking behavior within this critical period. For the subcategorized verbs, the majority of looks were in the direction of the appropriate picture. There was a significant main effect of verb class, ($F(2,24)=6.11$, $p<0.01$), and planned comparisons showed significant differences between the *VAP*-only and *VPA_ka*–only classes ($t=2.81$, $df=13$, $p<0.05$) and *VPA_ka*-only and alternating classes ($t=2.56$, $df=13$, $p<0.05$). However, alternating verbs also showed a *VAP* bias and the difference between alternating and one-construction verbs was not significant ($t=1.14$,

df=13, *p*>0.05), due to the strong bias towards the *VAP* construction for the alternating class, which was itself significant (*t=2.68, p*<0.05, *df*=13).

We also examined the data from one-construction verbs for effects of frequency. Although these effects were in the expected direction, they were not significant (frequency effect for *VAP* class: $F(1,12)=1.39$, $p>0.05$, frequency effect for *VPA_ka* class: $F(1,12)=0.26$, $p>0.05$, overall frequency effect: $F(1,12)=3.56$, $p>0.05$).

Discussion

The central finding of this experiment is that participants were able to acquire verb-specific subcategorization constraints from the input distribution. That is, they successfully learned the permissible and non-permissible constructions for each verb. Since there are no semantic or phonological cues to verb class and no pre-empting structures indicating missing ungrammatical forms, we conclude that this is accomplished through entirely statistical processes: tracking the frequency of the various verb-structure co-occurrences and using this as evidence as to the status of both the occurring and the missing verb-structure combinations. It is important to emphasize that these constraints were successfully learned even though the language showed heterogeneity in the way constructions were used with different classes of verbs. Although the majority of verbs in the miniature language were subcategorized to occur with only one construction, a substantial number could occur with both constructions (four verbs in each of the three classes). In addition, participants' ability to avoid over-generalization is particularly notable given that all test items were novel sentences (novel combinations of verbs and

nouns), so the learning could not have been accomplished by memorizing specific sentences or verb-noun combinations from the input exposure set.

Results from the three different tests were highly consistent: learners avoided overgeneralizations in their own productions, judged over-generalized forms to be less grammatical, and used their knowledge of verb subcategorization in real-time parsing. This last result was particularly striking in that we saw an influence of verb-class on eye-movements at the earliest possible moment, suggesting that participants were forming expectations about upcoming structure as soon as the verb was processed. This provides strong evidence that our artificial language stimuli are being parsed in a manner akin to natural language and establishes that purely distributional verb-biases can play a role in this process. The only difference between the three tests was that, whereas participants were clearly biased to predict the *VAP* construction and show a marginal preference for that construction in production, this bias did not affect judgments of grammaticality. (We return to this point in the Discussion of experiment 2).

In addition, we saw frequency effects within the one-construction verb classes: participants were less likely to produce or accept new structures with high frequency verbs. This result is important for two reasons. First, it mimics the entrenchment effects reported in the acquisition and processing literatures, suggesting that our learners are following a similar development path to learners of natural languages. (We acknowledge that, as our two constructions have the same meaning, our effects might also be described as *statistical pre-emption (*Goldberg, 2005) Whether the same effects would have been obtained with two semantically distinct constructions is an interesting question for future

research.) Secondly, observing learners' behavior with verbs for which they have had different amounts of exposure has provided us with a window into the learning process. It shows us the process by which our learners are able to retreat from over-generalization: the more often they hear a verb used in only one of the constructions, the more certain they are that it shouldn't be generalized. We expect that increased exposure to our languages would make learners increasingly less likely to over-generalize.

Experiment 2

The central finding of Experiment 1 was that learners were able to learn that certain verbs were (arbitrarily) subcategorized to occur only with one construction, even in the presence of a class of verbs which could occur in both. This constrained behavior can be described as *lexical conservatism* – the treatment of each verb was dependent on that verb's own behavior in the past. However, though we have emphasized this conservative behavior, which shows successful acquisition of the verb classes, our learners did sometimes generalize, as we saw whenever they made an 'error' with the one-construction verbs. In this case, the treatment of these verbs must be affected by the behavior of the *other* verbs in the input. Thus, as in real acquisition, learners of these languages must balance what they know about verbs in general against their verb-specific knowledge. Our hypothesis is that this balancing process depends on the *distributional details* on the input (though we acknowledge that other semantic and phonological factors will also play a role in natural languages). We have already seen one distributional factor which affects the tendency to generalize: individual verb frequency. In Experiment 2, we investigate whether the tendency to generalize a construction to a new verb might depend

on distributional factors *outside* of that verb's own history. In particular, we explore whether changing aspects of the class of alternating verbs might make learners more likely to extend that class and treat all verbs as alternating.

In Experiment 1, the presence of a class of unrestricted, alternating verbs did not prevent learners from acquiring subcategorization restrictions on other verbs. However, two aspects of the language may have helped to make the limits of that class minimally confusing. First, there were only half as many alternating verbs as one-construction verbs, so that learners observed mostly lexically conservative behavior. This may have encouraged them to attend to lexically-based distributional information. Second, although we included frequency distinctions within the one-construction verb classes, all verbs in the alternating class occurred equally frequently. This may have provided the learner with additional evidence that there was a distinct set of verbs which could occur with both constructions. If, in contrast, verbs within the alternating class occurred with varying frequencies, a rational learner might be less sure whether *all* verbs in the language might likewise fall in this class, since the absence of sentences in which some verbs occurred in certain structures might simply be due to the low frequency of those combinations.

In Experiment 2, we manipulated these factors. Learners were exposed to a language in which verb-structure co-occurrence patterns were identical to those in Experiment 1 (one-construction verbs occurred 100% of the time in one structure, alternating verbs occurred 50% of the time in each structure), but the class of alternating verbs had two new properties: It was twice as large (in terms of number of verbs and overall frequency) as the two one-construction classes together, and it contained verbs

which occurred with varying degrees of frequency. In addition, the frequency distinction within the two one-construction verb classes was increased from 3:1 to 5:1. This was done in an effort to increase the likelihood of finding these effects, particularly in the eye-tracking data, where they did not appear in Experiment 1.

Method

Participants

Fourteen native English speakers participated in the experiment. Most were undergraduate students at the University of Rochester; the remainder were former students or age mates who attended another college. Again, participants were run and tested individually, paid daily for their participation, and received a bonus upon completion of the entire experiment.

Description of the Language

The language used in Experiment 2 was identical to that used in Experiment 1 except for the division of verbs into verb classes.

Verb Classes: The 12 verbs were divided into *VAP*-only verbs, *VPA_ka*-only verbs and alternating verbs. However, in this experiment the distribution was as follows:

- 2 *VAP*-only verbs (1 high frequency, 1 low frequency)
- 2 *VPA_ka*-only verbs (1 high frequency, 1 low frequency)
- 8 alternating verbs (2 low frequency, 4 mid frequency, 2 high frequency)

Again, a different random assignment of verbs to verb classes was performed for each participant.

Input Sentence Sample: The input set in this experiment had the following distributional properties:

- Within the verb classes, verbs occurred equally often, except that high frequency verbs were five times as frequent as low frequency verbs and mid frequency verbs were three times as frequent as low frequency verbs.
- Overall, the *VAP*-only and *VPA_ka*-only constructions occurred equally often.

Materials

These were same as those used in Experiment 1.

Procedure

The procedure was very similar to that used in Experiment 1. In order to maintain the required frequency contrast (for some verbs to be heard five times as often as others), the exposure set was slightly larger. Also, no participants took tests on days 1 and 3 since that was found to have no effect on later performance in Experiment 1.

Procedure Days 1 to 3:

Part 1 - Vocabulary Learning: As in Experiment 1.

Part 2 - Sentence Exposure: Participants viewed a set of 216 transitive scene sentence pairs. The input set had the structure described above, with sentences presented in random order. Again participants were instructed to repeat the sentence out loud to help themselves learn.

Part 3 - Short Tests: (Day 2 only) As in Experiment 1.

Procedure on Days 4 and 5:

Part 1 - Vocabulary Test: As in Experiment 1.

Part 2 - Intermittent Short Exposure and Longer Tests.

Short Sentence Exposure 1: Participants viewed a subset of 108 sentences from the exposure set used on the previous days. Again they were instructed to repeat the sentence out loud to help themselves learn.

Production Test*:* As in Experiment 1.

Short Sentence Exposure 2: Participants viewed a subset of 36 sentences.

Eye-tracked Comprehension Test: As in Experiment 1.

Short Sentence Exposure 3: Participants viewed a subset of 36 sentences.

Grammaticality Judgment Test: As in Experiment 1.

Results

Production Test

Overall, 94% of participants' productions were found to conform to one of the two correct word orders. Again, incorrect productions were excluded from subsequent analyses. Figure 6 shows the proportion of trials in which participants produced the *VPA_ka* construction for each of the different verb types. As in Experiment 1, there is a main effect of verb class ($F(2,26)=4.38$, $p<0.05$), showing some learning of the verb subcategorization. However, the tendency to over-generalize is clearly stronger than in the previous experiment. This is particularly clear with the *VPA_ka*-only verbs, which are actually produced slightly more often with the *VAP* construction. In general, there is a strong bias for that construction (across the three verb classes, an average of 68% of productions occurred in that construction, compared to chance: $t=3.74$, $df=13$, $p<0.05$). There are no effects of verb frequency for one-construction verbs ($F(1,12)=.45$, $p>0.05$).

FIGURE 6 ABOUT HERE

Grammaticality Judgment Test

Difference scores are shown in Figure 7, with the raw judgment data for verbs in each construction shown in Table 2. In contrast to Experiment 1, the majority of verb-construction combinations were judged to be grammatical, including overgeneralizations with one-construction verbs. There is only a marginal effect of verb class ($F(1.09, 14.17)=3.90$, $p<0.1$, Greenhouse Geiser adjustment for sphericity). As in Experiment 1,

grammaticality judgments do not reflect the *VAP* bias found in production and processing (t=-0.22, df=13, p>0.05). There is no significant effect of frequency for one-construction verbs $F(1,13)=0.03$, $p>0.05$).

FIGURE 7 ABOUT HERE

TABLE 2 ABOUT HERE

On-line Comprehension Test

FIGURE 8 ABOUT HERE

Figure 8 shows the percentage of looks to the *VAP* picture in the critical ambiguous period, for each of the three classes of verbs. As with the production data, we see a bias for the *VAP* construction (across the three trials types, an average of 57% of looks were directed towards the *VAP* picture, compared to chance: t=3.16, df=13, $p<0.05$). Beyond this, there are no significant effects of verb type or verb frequency (verb type: $F(2,26)=0.69$, $p>0.05$; frequency: $F(1,11)=1.34$, $p>0.05$).

In sum, data from each of the three langauge tests provides only weak evidence that participants have acquired the verb subcategorization restrictions; there is clearly a strong tendency to generalize all verbs to both constructions. This is particularly clear in the grammaticality judgement data, where over-generalizations involving both constructions were judged to be grammatical the majority of the time. In the production and eye-tracking data, the bias for the *VAP* construction means that we largely see overgeneralization with *VPA_ka* only verbs.

One suprising aspect of these data is the lack of frequency effects, especially given that the frequency distinction was increased in this experiment. Interestingly, an analysis of data from the five participants who showed the best learning of verb subcategorization did reveal a significant effect of frequency in each of the three tests. This suggests the possibility that increased exposure may only play a role as learners become more certain that verbs are subcategorized.

Discussion

As predicted, learners in this experiment were more likely to generalize and use the one-construction verbs with the construction with which they had not appeared in the input (although, as in Experiment 1, the results from Production and Comprehension are complicated by a bias for the *VAP* construction, so that overgeneralization is more apparent with the *VPA_ka only* verbs).

One notable aspect of these results is the strong *VAP* bias in production and comprehension. Interestingly, as in Experiment 1, we see this bias only in production and processing, not in grammaticality judgments. Since there is no *VAP* bias in the input, why participants exhibit this preference is unclear. One possibility[3] is that we are seeing a transfer effect from the dominant SVO active form in English. However, agent-first structures are also more common cross-linguistically, suggesting that a preference for this form may be based on underlying cognitive or perceptual factors. Regardless of the explanation, it is of interest that in the more active tasks of production and real-time processing, linguistic behavior reflects a combination of the information acquired from the input and the inherent biases brought to the learning task. However, the fact that the

VAP bias does not affect the grammaticality judgments suggests that participants are able to isolate their language-specific grammatical knowledge.

However, the critical result in this experiment is the very different treatment of one-construction verbs as compared with Experiment 1. This is surprising, given the fact that individual verb-structure co-occurrence statistics are the same across the two languages: in *both* languages, any particular one-construction verb occurred 100% of the time with one construction. It is particularly striking that lexical constraints are more likely to be acquired even for low frequency verbs in Experiment 1 than for high frequency verbs in Experiment 2, although the latter have been heard five times more often in the input, and always in one construction. Thus it seems that in this language, the presence of the large alternating verb class provided evidence for generalization which outweighed evidence of lexically specific behavior, even for high frequency verbs.

Experiment 3

The findings of the first two experiments can be summarized as follows: Language learners can acquire arbitrary, verb-specific constraints. At the same time, this lexical information is in competition with the tendency to generalize. This competition can be influenced by manipulating the distributional details of the input, including both the frequency of the individual verbs and the relative size of the alternating verb class. In Experiment 3 we aimed to further investigate the nature of both lexically-based and more general learning, and to further probe the balance between these two processes.

One important feature of the languages used in Experiments 1 and 2 was that the lexical subcategorization constraints were always absolute in nature. That is, each of the subcategorized verbs occurred in the same construction 100% of the time. In contrast, alternating verbs were unconstrained in that they occurred equally freely in each construction (although the agent-first bias did impose an unexpected constraint against the *VPA_ka* construction). The languages were so set up in an attempt to mimic the dichotomous pattern which has often been assumed in the acquisition literature. However, as discussed in the introduction, natural language verbs may be more probabilistically constrained, and these statistics may also influence both online processing (e.g. Trueswell et al., 1993) and production probabilities during learning (Wonnacott & Newport, 2005). In addition, learners may also acquire probabilistic constraints *above* the level of particular verbs (Juliano & Tanenhaus, 1994). For example, learners may acquire and utilize the language-wide likelihood of encountering each of the constructions in the input, independent of the verb with which it occurred. In our previous experiments it was not possible to pull apart the influence of this higher level statistic from chance behavior, since our two constructions occurred equally often across the verbs of the language. Thus although we know that the treatment of a verb can be influenced by the behavior of other verbs in the input (or our learners would never produce overgeneralizations), we don't know how precise that generalization process is.

In Experiment 3, we address whether our learners can acquire statistical biases at both the verb-specific and verb-general levels, and how these statistics affect each of the three language behaviors in our paradigm. In addition, we were again interested in exploring how the tendency to generalize (and potentially apply verb-general patterns)

was influenced by the way in which constructions were distributed across verbs. In Experiments 1 and 2, we contrasted languages that differed in how many verbs in the language exhibited alternating constructions. In Experiment 3 we compared learning of two extreme languages: one with the largest possible alternating class (all verbs alternate), and one with the smallest possible alternating class (no verbs alternate). The two languages were designed to contain the following verbs, all equally frequent:

Language 1 - *Lexical* language*:* 1 *VAP*-only verb, 7 *VPA_ka only* verbs

Language 2 - *Generalist* language*:* 8 *VPA_ka-biased*, alternating verbs (each occurred seven times more often in a *VPA_ka* sentence than in a *VAP* sentence).

This design will allow us to compare the learning of absolute lexical constraints and probabilistic lexical constraints, by comparing one-construction verbs in the lexical language versus biased but alternating verbs in the generalist language. We can also ask whether learners can acquire and use verb-general statistics. In each of the two languages, there is a 7:1 bias for the *VPA_ka* construction. We can therefore ask whether learners generalize this bias to their use of new verbs. Perhaps most important, we will explore the extent to which generalization is dependent on the pervasiveness of alternation across the language. The absolute and relative frequencies of the two constructions, *VAP* and *VPA_ka*, are matched across the two languages; but the two languages are extremely different in how these constructions are distributed within individual verbs. As compared to the previous experiments, will learners be much more likely to generalize when exposed to the *generalist* language, and much less likely to generalize when exposed to the *lexicalist* language?

One addition to this experiment was that we introduced the use of *novel verbs* into testing. (In previous experiments, while sentences were novel, the nouns and verbs used in test items were the same nouns and verbs used throughout the exposure set.) Our initial aim was to explore how learners would generalize from the input to verbs which had no distributional history. Since our design allowed the use of four novel verbs, we adopted a between-participants design for this test types: one group of participants saw these verbs in the *Production* test, one group saw them in the *Grammaticality Judgments*, and one group saw them in the *Online Comprehension* test. In addition, we were also interested in seeing how learners would generalize when given even a small amount of evidence linking the verb to one of the two constructions. (Note that this type of generalization across constructions is what we have seen with the one-construction verbs in Experiments 1 and 2). In order to look at both types of generalization within the same experiment, while at the same time making maximum use of resources, we decided to capitalize on the fact that taking the comprehension test provides learners with further exposure to the language. That is, when a verb occurs within a test sentence in test, the learner is hearing that verb used in one of the two constructions. With this is in mind, we deliberately set up the comprehension test so there were four test sentences for each of the new verbs and, crucially, each of these test sentences involved the same construction. Thus taking the comprehension test provided a small amount of lexically consistent exposure to each of the new verbs (so that they became like very low frequency one-construction verbs). In order to see the influence of this exposure when the new verbs occurred in later tests, we adopted a design in which the same new verbs were used across the different tests, but different groups of participants received the tests in a

different order. Thus for each of the three tests there was a group of participants who encountered the new verbs for the first time in that test (each of these verbs was then considered novel for that group, except in the case of the comprehension test where only the first instance of each verb in the test could be considered novel), and there was a group of participants who had previously encountered these verbs in the comprehension test (in the case of the comprehension test itself, we took this to be the third and fourth instance of that verb in the test). Thus we could explore the treatment of entirely novel verbs, and also verbs for which participants had a very small amount of lexically consistent exposure.

Method

Participants

Thirty native English speakers participated in the experiment, all undergraduate students at the University of Rochester. Participants were randomly divided into two conditions, each exposed to one of two different input languages. Participants in each of these conditions were further subdivided into three sub-groups who each took the tests in a different order on day 5 (see *Procedure*).

Description of the Two Languages

Both of the languages in this experiment used the same vocabulary and basic grammatical constructions as the previous experiments. Again what was manipulated was the pattern of verb-structure co-occurrences.

Lexical Language*:* 1 *VAP*-only verb, 7 *VPA_ka only* verbs

Generalist Language*:* 8 *VPA_ka-biased,* alternating *verbs* (each occurred seven times more often in a *VPA_ka* sentence than in a *VAP* sentence).

All verbs in each of the languages were equally frequent, so that the frequency of the two constructions was matched across the two languages (i.e., the *VPA_ka* construction is 7 times more frequent than the *VAP* construction in both languages).

Note that each of these languages contained only 8 verbs, compared to 12 verbs in Experiments 1 and 2. This was done so that 4 verbs could be reserved for use as novel verbs in the tests. As in the previous experiments, particular verbs were randomly assigned to different verb types for different participants. Thus new and familiar verbs were different subsets of the 12 available verbs for different individuals, and there were no systematic semantic or phonological relationships between the two.

Materials

The same as those used in Experiments 1 and 2.

Procedure

The procedure was identical for each of the two language conditions and was very similar to that in the previous experiments. However, the inclusion of new verbs in testing (on day 5 only) involved introducing some more substantial changes in procedure on that day.

Procedure Days 1 to 3.

All procedures and instructions were identical to those in Experiments 1 and 2, unless otherwise stated.

Part 1 - Vocabulary Learning.

Part 2 - Sentence Exposure: Participants viewed a set of 128 transitive scene sentence pairs from the relevant language (8 verbs * 16 instances apiece).

Part 3 (on day 2 only) - Short Tests (only familiar verbs test items): As in previous experiments, except that in each test there were 2 test items per verb (making 16 rather than 12 test items in total). In each condition, the set of test items was structured so as to affect verb-general statistics as little as possible. In the *lexical* condition, the two test items involving the *VAP*-only verb were *VAP* sentences and the other fourteen test items were *VPA_ka* sentences. In the *generalist* condition, two test items involving four different verbs were *VAP* sentences, and all other test sentences were *VPA_ka* sentences. As before, these tests were included to help participants learn.

Procedure on Day 4.

Part 1 - Vocabulary Test

Part 2 - Intermittent Short Exposure and Longer Tests (only familiar verb test items).

Short Sentence Exposure 1: Participants viewed a subset of 64 scene sentence pairs (8 instances of each verb) from the exposure set used on the previous days. In each

case, this subset maintained the statistical structure of the entire input language, both lexically and across verbs.

Production Test: As on day 2, except that there were 4 test items for each familiar verb, 32 in total, and none of the specific scenes used in the test had been seen in the input. Test items were scored as in Experiments 1 and 2.

Short Sentence Exposure 2. Participants viewed another subset of 32 scene sentence pairs (4 instances of each verb).

On-line Comprehension Test: As on day 2, except that there were 4 test items for each verb, 32 in total. In the *lexical* condition, the four test items involving the *VAP*-only verb were *VAP* sentences, and all other test items were *VPA_ka* sentences. In the *generalist* condition, four test items involving four different verbs were *VAP* sentences, and all other test sentences were *VPA_ka* sentences. Also, in contrast with day 2, none of the specific sentences used had occurred in the input, and the test was now eye-tracked (all eye-tracking procedures as in Experiments 1 and 2).

Sentence Exposure 3: Participants viewed another subset of 32 scene sentence pairs (4 instances of each verb). Test items were scored as in Experiments 1 and 2.

Grammaticality Judgment and Ratings Test: As on day 2, except that there were 4 test items for each verb: 2 in *VAP* sentences and 2 in *VPA_ka* sentences (32 test items in total), and none of the sentences used had occurred in the input. Test items were scored as in Experiments 1 and 2.

Procedure on Day 5:

Part 1 - Vocabulary Test

Part 2 - Intermittent Short Exposure and Longer Tests (Including Novel verb Test Items)

Short Sentence Exposure 1: As on day 4.

Test 1: Either Production, Comprehension or Grammaticality Judgment.

Short Sentence Exposure 2: As on day 4.

Test 2: Either Production, Comprehension or Grammaticality Judgment.

Short Sentence Exposure 3: As on day 4.

Test 3: Either Production, Comprehension or Grammaticality Judgment.

Details of Day 5 tests:

Each of the three different tests on day 5 included 16 familiar verb test items (8 familiar verbs * 2 test items each) and 16 test items which included the 4 'new' verbs reserved from the input set (4 new verbs * 4 test items each). The new verb test items involved the same five nouns which had occurred in sentence exposure[4] . New and familiar verb test items were randomly intermixed, though in each case the first 8 test items involved familiar verbs. Participants were warned before each test that they might be given test items containing actions and verbs which they had not encountered during

learning, but were given no other special instructions for dealing with these new verb test items, which were exactly like familiar verb test items. For example, in the Production test, participants would see a scene containing a new action, hear one of the novel verbs, and complete the sentence. In the Grammaticality Judgment and Comprehension tests, participants heard complete sentences involving the novel verbs and saw appropriate scenes with new actions. In the grammaticality judgment test, each verb occurred twice with each of the constructions. In the Eye-tracked Comprehension Test, two of the verbs were always heard in the *VAP* sentences (that is, all 4 test items with that verb used that construction), and two were always heard in *VPA_ka* sentences. This test thus provided learners with a small amount of lexically specific input – a fact that we aimed to exploit.

Since we were interested in seeing how participants performed with new verb test items both when the verbs were entirely novel and when they had the previous exposure provided by the comprehension test, the 15 participants in each language condition were further (randomly) divided into 3 three sub-groups which each took the tests in different order:

Group 1: *Comprehension* → *Production* → *Grammaticality Judgment*

Group 2: *Production* → *Grammaticality Judgment* → *Comprehension*

Group 3*: Grammaticality Judgment* → *Comprehension* → *Production*

Note that the new verbs could only be considered truly novel in the *first* test for each group, i.e. Production for Group 2, Grammaticality Judgment for Group 3, Comprehension Group 1 (though in this last test only the *first* test item for each verb

could actually be considered novel). It is these data that were included in the *Novel Verbs* results below.

In addition, to provide preliminary data about the effects of a small amount of consistent lexical input, we considered the treatment of the new verbs by participants in Group 3, who took the Comprehension test before both of the other tests. For the Comprehension Test itself, we considered only the *third* and *fourth* test items for each verb. For these participants, these new verbs are not novel, and these data are thus included in the *Minimal Exposure Verbs* results below.

New verbs test items were also included in the later day 5 tests given to Groups 2 and 3, that is, in the Grammaticality Judgment and Comprehension tests for Group 2 participants, and in the Comprehension and Production tests for Group 3 participants. However, these data were *not* included in the analyses below. The test items were included so that all groups of participants took the same sets of tests and the relevant *familiar verbs* data could be collected.

Results

Production Test

Overall, 95% of productions in each language conformed to one of the two permissible words orders, showing again that participants had mastered the basic constructions in each language. Figure 9 shows, for each verb type in each of the language conditions, the proportion of productions which conformed to the *VPA_ka* construction. (In this experiment we chose to score with respect to that construction,

since that was the dominant construction for this experiment.) We discuss data from each of the three verb types in turn.

FIGURE 9 ABOUT HERE

Familiar Verbs. Production probabilities closely match those of the input languages. In the *lexical* condition, participants strongly distinguish the two verb classes, primarily using *VAP*-only verbs in the *VAP construction* and always using *VPA_ka*-only verbs in the *VPA_ka* construction, with very few overgeneralizations. In the *generalist* language, productions approximately matched the probabilities of the input: across verbs, 86% of productions used the *VPA_ka* word order. A between-participants comparison of production probabilities with the *VPA_ka*-only verbs in the *lexical* condition and with the biased alternating verbs in the *generalist* condition showed a marginally significant difference (t=2.02, p<0.06, df=28). This provides evidence that production probabilities may mirror the *degree* of verb bias in the input.

Novel verbs. These data were collected from the group of participants in each condition who saw these verbs for the first time in this test (participant Group 2)[5]. It is clear that, in both language conditions, productions with novel verbs demonstrate a preference for the *VPA_ka* form (*lexical* t=11.7, p<0.001, df=4; *generalist* t=4, p<0.02, df=4), and there is no significant difference between the two conditions (t=0.78, p>005, df=4). Since, in both conditions, participants had received no input for the novel verbs, we conclude that the bias for the *VPA_ka* construction had been abstracted by observing the usage of constructions across the language as a whole.

Minimal Exposure Verbs. These data were collected from the group of participants who took the Production test *after* the comprehension test on day 5 (Group 1 in each language condition) [6]. As described above, each of these minimal exposure verbs had previously been heard in the comprehension test. Two of the four verbs had been heard four times in the *VAP* construction, and two had been heard four times in the *VPA_ka* construction. We refer to these subsets as *'VAP* minimal exposure verbs' and *'VPA_ka* minimal exposure verbs,' respectively. (In fact, they could be considered to be very low frequency *VAP*-only and *VPA_ka*-only verbs).

Recall that conclusions from the minimal exposure conditions are necessarily qualified by the small number of participants, due the fact that we can only observe effects of minimal exposure for Group 1 participants. With this caveat in mind, minimal exposure to one of the two constructions seems to be treated strikingly differently by learners exposed to the two different languages. In the *lexical* condition, participants clearly distinguish two classes of verbs (t=4.09, p<0.01, df=4). These verb types appear to be treated much like *VAP*-only and *VPA_ka*-only familiar verbs; that is, they are primarily produced with the constructions in which they were heard in the four exposures of the comprehension test. In contrast, in the *generalist* condition we see no effect of the specific restricted constructions used in the exposures for these verbs. In this case, as with entirely novel verbs, participants display a bias for the *VPA_ka* construction for both classes of verbs, matching the verb-general bias that holds across the language, rather than the limitation to the construction in which they had heard the verbs.

Grammaticality Judgments

Figure 10 shows the grammaticality judgment data for the different verb types in the different languages. Again we show difference scores, here scored with respect to the *VPA_ka* construction; the raw judgment scores from which these were calculated are shown in Table 3.

FIGURE 10 ABOUT HERE

TABLE 3 ABOUT HERE

Familiar verbs. First, regarding the *lexical* language, it is again clear that grammaticality judgments strongly reflect absolute grammatical restrictions: *VAP*-only verbs are more often judged grammatical in the *VAP* construction than in the *VPA_ka*, (t=8.5,$p<0.0001$,df=14), whereas *VPA_ka*-only verbs are judged more grammatical in the *VPA_ka* construction, (t=7.9,$p<0.0001$,df=14).

For the *generalist* language, again the results are quite different; both constructions are accepted as grammatical. There is a weak bias for the *VPA_ka* form; however, this difference is only marginally significant (t=2.0, $p<0.1$, df=14). In fact, over 80% of the non-dominant *VAP* sentences are accepted as grammatical. This contrasts strongly with data from the Production test, where we saw that *VAP* sentences constituted less than 15% of participants' own productions. An examination of individual participants' response patterns was particularly revealing: more than half of participants accepted the two sentence types equally often, despite using the *VPA_ka* form far more

often in production. Thus probabilistic lexical constraints do not seem to have a strong influence on grammaticality judgments.

Novel verbs. (These data were collected from the group of participants in each condition who saw these verbs for the first time in this test: participant Group 3.) In the generalist condition, participants again accept the majority of sentences, and are no more likely to accept one sentence type than the other (t=1, p>0.05, df=4). Thus we see no influence of the general bias for the *VPA_ka* form that was reflected in production probabilities. We do see some *VPA_ka* bias in the lexical condition (t=1.8, p>0.05, df=4). Although this is not significant, this could be due to lack of power. We therefore repeated the test including the results from Group 1 participants (participants who had previously seen the new verbs test in the production test but not the comprehension test), since production did not provide participants with any input. This yielded a significant result (t=2.6, p<0.05, df=10). These differing results from the two conditions are puzzling: why should input statistics influence the treatment of novel verbs only in the lexical language? If anything, we would have expected the reverse pattern, that participants would only generalize a statistical pattern which holds across verbs when that pattern also holds at the level of individual verbs. An examination of the treatment of individual verbs by individual participants yielded an explanation. In every case, an individual verb is judged either to be always grammatical in both construction (as in the generalist condition) or to be only grammatical in the *VPA_ka* construction. This latter treatment matches that of the *VPA_ka*-only verbs in the input and suggests that in this case participants are assuming those verbs to be members of that class.

Minimal exposure verbs. (These data were collected from the group of participants who took the Production test *after* the comprehension test on day 5- Group 1 in each language condition).

Again we see that the small amount of lexical exposure provided for these verbs has had very different effects for learners in each of the two languages. In the *lexical* condition, participants strongly distinguish the two verb classes (t=15.7, p<0.0001,df=4). In contrast, there are *no* effects of lexical exposure in the *generalist* condition; participants judge these verbs to be equally grammatical in the two constructions, just as with the novel verbs in this condition. Thus as in production, minimal exposure verbs are treated just like *VAP*-only and *VPA_ka*-only verbs in the *lexical* condition, and like novel verbs in the *generalist* condition.

Eye-tracked Comprehension Test

FIGURE 11 AROUND HERE

Familiar verbs. Figure 11 shows eye-tracking data with familiar verbs in both language conditions. In the *lexical* condition, for both *VAP_only* and *VPA_ka* only verbs, looks to the picture corresponding to the sentence structure begin to rise as soon as participants are able to identify the first noun. Thus it is clear that, as in the previous experiments, knowledge of the verb constraints direct looking preferences from the earliest possible moment. In the *generalist* condition, all sentences begin with a *VPA_ka* biased alternating verb, and we see looks to the *VPA_ka* rise during the first noun, indicating that they anticipate that construction. Comparing looking behavior with

VPA_ka-only verbs (in the *lexical* condition) and biased alternating verbs (*generalist* condition), the bias to look at the *VPA_ka* picture emerges a little earlier for the absolutely constrained verbs (around 700ms) than for the statistically constrained verbs (around 900ms). In order to directly compare looking behavior across conditions, Figure 12 shows the percentage of looks directed at the *VPA_ka* picture, out of total looks to each picture, during the critical period of the sentence (between the beginning of the first noun and the 'ka' particle). The three different verb types are clearly differentiated. Importantly, there is a significant difference in the proportion of looks directed at the *VPA_ka* with the *VPA_ka*-only and biased alternating verbs ($t=2.6$, $p<0.02$, $df=25.5$, adjusted due to heterogeneity of variance), providing evidence that real time processing is sensitive to the degree of verb bias in the input.

FIGURE 12 AROUND HERE

Novel verbs. Since the comprehension test itself provides minimal exposure input, the only truly novel verbs in testing are the first instances of each new verb heard in the test. Unfortunately, it proved impossible to collect interpretable eye-movement data from these test items, since this is also the first time participants encounter the visual action represented by the verb.

Minimal Exposure Verbs. Figure 13 show the pattern of looks to each picture with minimal exposure *VPA_ka* and minimal exposure *VAP* verbs. These data were collected from the group of participants who took the Online Comprehension Test first on day 5 and only include the *third* and *fourth* instance that each of the new verbs appeared in the test. Thus minimal exposure *VAP* verbs are verbs which had been heard twice in

the *VAP* construction; minimal exposure *VPA_ka* verbs are verbs which had been heard in twice in the *VPA_ka* construction.

FIGURE 13 ABOUT HERE

Although these data are based on relatively few trials (5 participants * 2 trials per participant for each verb type), the pattern of looks is nevertheless suggestive. In the *lexical* condition, for the minimal exposure *VAP* verbs participants begin to show a preference to look at the *VAP* rather than *VPA_ka* picture at around 720 ms, as soon as the first noun is identified. Similarly, for the minimal exposure *VPA_ka* verbs, they show a looking preference for the *VPA_ka* picture at around 800 ms. Thus these verbs appear to be treated much like *VAP*-only and *VPA_ka*-only verbs. In the *generalist* condition, the pattern is less clear. With the minimal exposure *VAP*-only verbs there are, by chance, more looks to the *VAP* picture at the beginning of the sentence (even in the period before the first noun was heard, indicating random behavior). However, we do see looks to the incorrect but biased *VPA_ka* construction rising from around 1200ms. This suggests that the verb-general bias for that form is influencing looking behavior. Similarly, we see looks to the *VPA_ka* construction rising from around 1040ms with the minimal exposure *VPA_ka* verbs. Thus these data suggest a similar pattern of results to that seen in production: in the *lexical* condition, participants rely on the minimal amount of verb-specific information they have received, whereas in the *generalist* condition they are influenced by the verb-general patterns. However, there were insufficient data to perform statistical analyses.

Discussion

This experiment yielded a large amount of data, and the results have a number of theoretical consequences. Below we summarize the results obtained with each of the three verb types across the three language tests.

Familiar verbs: These data demonstrate that, as with real languages, learners of artificial languages are able to acquire both absolute and probabilistic subcategorization constraints. In this experiment we saw that these statistics affect both production probabilities and expectations in online processing. Moreover, both of these behaviors reflect the degree of statistical preference exhibited by a particular verb for a particular structure (i.e., the contrast between a 100% versus and an 87% bias for a particular structure). Interestingly, though grammaticality judgments were strongly affected by absolute subcategorization, they were only weakly affected by the statistical bias.

Novel Verbs (and minimal exposure verbs in generalist condition): These data demonstrate that learners are able to acquire statistical information *above* the level of particular verbs, what we refer to as *verb-general* statistics. Again these statistics are reflected in production probabilities and in expectations during online processing. The ability to acquire this bias is particularly striking, given that it must be working against participants' inherent preference for the *VAP* construction, which was seen in Experiments 1 and 2. Again it is interesting that very different results were seen with grammaticality judgments. Here the verb-general statistics had no influence; participants judged these verbs to be equally grammatical in each of the two structures.

Minimal exposure verbs (in both language conditions): These verbs are treated very differently by learners in the two language conditions. This is striking, given that each group of learners has heard these verbs used in exactly the same set of sentences. What differs for the two groups is their experience with *other* verbs in the language. Learners in the *generalist* language, learning in an environment in which all verbs are biased alternating, treated the minimal exposure verbs as though they also conformed to that pattern – ignoring the small amount of verb-specific exposure they received. In contrast, learners of the *lexicalist* language, learning a language with no alternation, relied on the four verb-specific exposures they received and, ignoring the verb-general pattern, treated these verbs like one-construction verbs. These results corroborate and extend the results of Experiment 2: increasing the size of the alternating class increases the tendency to generalize, even for verbs that have previously only occurred in one construction. Here we see that making the class sufficiently large or small can determine whether learners use verb-general or verb-specific patterns.

Overall, the results of this experiment show that learners can acquire verb-specific constraints (both absolute and probabilistic) and also verb-general constraints; but which of these types of statistics influences language behavior depends on the nature of the language being learned. In addition, the results show an interesting difference among the three types of language test: while *Production* and *Online Processing* data reflect both absolute and probabilistic constraints, grammaticality judgments primarily reflect only absolute subcategorization constraints. This suggests that participants use the grammatical judgment tasks to express a notion of 'grammaticality', which can be

disassociated from probabilistic information (although what is judged to be grammatical is a consequence of the statistics of the input).

General Discussion

This research explored whether adult participants could learn the distributional relationships between twelve verbs and two constructions in an artificial language. Across three experiments, learners were found to track both *verb-specific statistics* (that is, the likelihood of a particular verb co-occurring with a particular argument structure), and also *verb-general statistics* (that is, the likelihood of a particular argument structure occurring across the verbs of the language). Importantly, however, how learners utilized these competing sources of information depended upon additional statistical factors: *verb frequency* (learners were more likely to ignore verb-specific statistics with low frequency verbs) and the *distribution of verb types across the language* (learners were more likely to ignore verb-specific statistics in languages with a large alternating verb class).

These factors had subtly different effects for the three different language tests. Both production probabilities and the expectations used in on-line processing reflected either verb-specific statistics or, when this was unreliable, verb-general statistics. In contrast, grammaticality judgments were only strongly influenced by *absolute* verb-specific constraints. A verb was judged to be more grammatical in one construction than the other when the participants were convinced that that verb could *only* occur in that construction. There was little effect of differing probabilities of subcategorization and no effect of verb-general preferences on these judgments; when these probabilities varied, constructions were nonetheless judged equally grammatical with that verb. There was

also no effect of the general *VAP* bias that showed up in production and on-line comprehension in Experiments 1 and 2. (We point out, however, that it is *not* the case that participants simply accept combinations which they have heard and reject those that they have not.) These results suggest that our language users can distinguish grammatical knowledge from other factors that affect the usage of verbs and structures. This confirms a distinction that has commonly been assumed in linguistics. It also highlights the methodological importance of considering different types of linguistic behavior when assessing what has been learned during acquisition.

These results have implications for several central theoretical issues in language acquisition and language processing. First, our findings address Baker's Paradox regarding how learners are able to generalize yet avoid inappropriate overgeneralization. Our work suggests that the solution lies in the fact that learners are highly sensitive to the distributional details of their input[7], and that deciding whether to restrict the use of a construction to the particular verbs with which it has occurred in the input, versus generalize its use, is part of a larger process of appropriately balancing verb specific and verb general patterns of usage.

Second, our results speak to a longstanding issue in language acquisition regarding the roles of distributional and semantic cues in acquiring syntax (Bates & MacWhinney, 1982; Bowerman, 1978; Moeser & Bregman, 1972; Morgan & Newport, 1981). One possibility would be that learners should only acquire linguistic regularities that serve some functional or communicative purpose. The current experiments add to a body of artificial language learning experiments which show that this is *not* the case

(Gomez, 2002; Mintz, 2002; Morgan, et al., 1987; Morgan & Newport, 1981; Saffran et al., 1996). Of course, in natural language the interplay between verb semantics and verb distribution is particularly rich and complex, and we know that this information plays a role in adult language (see Bresnan, Cueni, Tatiana, & Baayen, 2005, for a demonstration of the complex set of mutually constraining influences - semantic, pragmatic and phonological - affecting the production of the two dative constructions in English). It seems likely that this same information will also play a role in acquisition, and indeed we reviewed experimental evidence demonstrating that children can make inferences as to whether a verb can occur in a particular structure from its meaning, and about its meaning given its occurrence in a particular structure (e.g. Gropen et al., 1991; Naigles 1990). On the other hand, work by Gleitman and colleagues suggests that the situation in which learners have access to verb meaning and not verb distribution may actually be very rare outside of the lab (Gillette et al., 1999; Snedeker, 2000). The current work demonstrates that the acquisition of verb syntax does not necessarily rely on the prior process of extrapolating verb meaning from environmental contingencies. It also provides direct evidence for an important assumption of the Syntactic Bootstrapping hypothesis, that learners are able to learn the range of structures in which a verb can occur prior to learning its meaning.

In future work it will be important to augment the artificial languages to contain various semantic cues, in order to address the acquisition of verb meaning, verb distribution, and the complex correspondences between the two. It seems likely that different semantic relationships may emerge at different stages of development. For example, the relationships between syntactic transitivity and causal events, and between

ditransitivity and events of transfer, appear to hold cross-linguistically (Fisher et al., 1991; Gleitman, 1990) and even in the invented gesture systems of linguistically deprived deaf children (Feldman, Goldin-Meadow, & Gleitman, 1978; Goldin-Meadow & Mylander, 1998), suggesting that learners may be predisposed toward some of these fundamental constraints (although it is not clear whether this is the result of innate linguistic knowledge, Lidz, Gleitman, & Gleitman, 2003, or of more general pragmatic principles, Goldberg, 2004). Thus we would expect these factors to play a role in (mutually) constraining verb-syntax from an early stage in learning. In contrast, more idiosyncratic semantic classes, such as Pinker's 'narrow' English verb classes, may emerge later in the learning process, only after a considerable amount of distributional learning has taken place, on the basis of the semantic attributes of sub-groups of verbs occurring in similar environments (see Goldberg, 1995 and the model presented in Allen, 1997 for accounts along these lines). In addition, exploring the acquisition of a pair of alternating argument structures which each have a different semantics (akin to the transitive-intransitive causative alternation in English) will also provide an opportunity to explore a question raised by Goldberg (2005): can the process of entrenchment occur in a situation where the usage of one structure does not provide direct evidence of the non-usage of the other[8]?

Our findings also highlight the importance of distributional learning for theories of sentence processing. Of course we do not suggest that semantic and pragmatic cues play a minimal role in natural language processing. There is abundant evidence that

comprehenders form expectations on the basis of multiple sources of information (see Tanenhaus & Trueswell, 1995 for a review), and specifically that knowledge of verb sense plays a role in determining a verb's subcategorization profile (Hare et al., 2003). Nevertheless, the current results clearly demonstrate that statistical effects can stand alone and that processing is sensitive to quite fine-grained statistical detail. In addition, we have demonstrated that learners are sensitive to statistical information *above* the level of individual verbs. This is consistent with input-driven accounts of non-lexical structural preferences (Juliano & Tanenhaus, 1994; Tabor et al., 1997; Kim et al., 2002), suggesting that the preference to interpret post-verbal nouns as direct objects in English may indeed be due to the predominance of that type of complement structure in the language.

Finally, our results provide support for the hypothesis that learners adopt a *rational* or *Bayesian* procedure for evaluating competing sources of statistical information. The problem we have investigated concerns how learners determine the usage of different argument structures with different verbs. That is, how do learners decide when a missing verb-construction combination is grammatical but happens not to have occurred in the sample of speech they have heard, and when it is ungrammatical in the language. Two potential sources of information are the past usage of those constructions with each particular verb, and the past usage of those constructions across the verbs of the language. The learner's task is to determine which of these provides the best predictor of the future usage of constructions with each verb. Across these three experiments, our learners appear to use a variety of statistics about the language to estimate whether verb-specific or verb-general statistics are more or less likely to reveal the underlying nature of verb usage. For example, if a verb is high in frequency, this

provides a large sample of verb-specific information, making it sensible to rely on that source of information when determining that verb's future behavior; if a verb is low frequency, verb-specific information should be less revealing of the appropriate restrictions for that verb. Similarly, in the languages we tested, increasing the size of the alternating verb class increases the number of individual verbs which adhere to the verb-general pattern, thus increasing the utility of that statistic for predicting the usage of any particular verb; in contrast, in a language with large numbers of idiosyncratic verbs, the best source of information about a verb is its own behavior. These findings concur with the Hierarchical Bayesian model presented by Kemp, Perfors, & Tenenbaum (in press). This model classifies objects according to their feature distributions by not only estimating the feature distribution of each *particular* object, but also two 'over-hypotheses' about the nature of the distribution: the overall distribution of features for each class, and the variability of individual objects within each class (how well the distribution for each object conforms to the overall distribution of the class). Ongoing research investigates whether this algorithm can account for the current data. By combining experimentation with this and other modeling techniques, we hope to clarify further the precise learning algorithms at work in these experiments. For example, what exactly is it about increasing the size of the alternating verb class that increases generalization? One possibility is that what is important is the overall frequency of alternation; however, it may be that the critical factor is the number of verbs in that class (this concurs with the notion of 'type' frequency, a factor known to be associated with productivity, see Bybee 1995; Goldberg, 1995; 2007- though here we are concerned with the type frequency of a shared distributional pattern, rather than of a particular linguistic

structure). There are situations in which these two possible explanations could lead to very different behaviors, for example if a language contained just one very-high frequency alternating verb. Another question is whether variety in the sense of frequency differences, a factor manipulated in Experiment 2, actually increases the tendency to generalize (in Experiment 2 the effects of this factor are confounded with the increase in the size of the class).

To fully evaluate the impact of our results for models of language acquisition, two further lines of research are necessary. First, it will be important to explore spoken corpora to ascertain how the factors identified in these experiments apply to different natural language constructions and how these statistical patterns line up with the varying degrees of generalization shown by native speakers for different structures. Second, it will be important to repeat these experiments with children, to determine if they apply the same learning algorithms as adults for these variables, given precisely the same input. There are a number of ways in which children and adults have been shown to differ as language learners, though they do behave similarly in many ways as well (Johnson & Newport, 1989; Newport, 1990). One way in which the behavior of our adult participants differs from that reported in the developmental literature is that our learners generalize in all of the three language behaviors tested. In contrast, there is evidence that young children are reluctant to produce verbs in new constructions, (e.g. Leiven, Pine, & Baldwin, 1997; Tomasello, 1992; Tomasello, Akhtar, Dodson, & Rekau, 1997), though they show a more generalized knowledge earlier in comprehension, particularly when looking time measures are used (Gertner et al. 2006; Naigles, 1990). It is possible that child learners will show similar differentiated behaviors across the tests in our

experiments, or perhaps they will show an initial tendency to be lexically conservative across measures (Tomasello, 2000). Of course children do eventually generalize, and when they do so, they also sometimes overgeneralize (Austin, Newport & Wonnacott, 2006; Ervin, 1964; Hudson Kam & Newport, 2005; Singleton & Newport, 2004). From the perspective of the current results, the most interesting questions will concern whether children retain a tendency to focus on lexically-based patterns once they have also begun to generalize – that is, do they give more weight to verb-general statistics than to verb-specific statistics, or (like adults) do they weigh these variables differentially and rationally, depending on the character of the language they are learning and the amount of input they have received? Ultimately, the critical issue will be understanding whether there are such differences between adult and child learners, and how any differences between child and adult learners may contribute to the greater success of children at acquiring natural languages.

References

Allen, J. (1997). Probabilistic constraints in acquisition. *Proceedings of the GALA conference on language acquisition*. Edinburgh: University of Edinburgh.

Ambridge, B., Pine, J.M., Rowland, C.F. & Young, C.R. (in press) The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition.*

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73(3), 247–264

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science, 9*, 321-324.

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

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

Bates, E., & MacWhinney, B. (Eds.) (1982). *Functionalist approaches to grammar*. Cambridge: Cambridge University Press.

Bowerman, M. (1978). Structural relationships in children's utterances: syntactic or semantic? In L. Bloom (Ed.), *Readings in Language Development*, New York: Wiley.

Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals.* (pp. 83-101): Oxford: Basil Blackwell.

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. Merriman, E. (Eds.), *Beyond names of things: young children's acquisition of verbs.* (pp. 353-376), Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

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.

Brooks, P. J. & Zizak, O. (2002). Does pre-emption help children learn verb transitivity? *Journal of Child Language*, 29, 759-781.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley.

Bybee, Joan. (1995). Regular Morphology and the Lexicon. *Language and Cognitive Processes* 10(5):425–55.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Clifton, C., Frazier, L., & Randall J. (1983), Filling gaps: Decision principles and structure in sentence comprehension", *Cognition,* 13(2), 187-222

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6* 84-107.

Matthews, D.E., Lieven, E.V.M., Theakston, A., Tomasello, M. (2005). The role of frequency in the acquisition of English word. *Cognitive Development, 20* 121-136.

Ervin, S. M. (1964): Imitation and Structural Change in Children's Language. In E. Lenneberg (ed.): New Directions in the Study of Language, 163-189. Cambridge, MA: MIT Press.

Feldman, H., Goldin-Meadow, S., & Gleitman, L. R. (1978). Beyond Herodotus: The creation of language by linguistically deprived deaf children. In A. Lock (Ed.), *Action, symbol, and gesture: The emergence of language.* New York: Academic Press.

Ferreira, F., & Henderson, J. M. (1990). The use of verb information in syntactic parsing: A comparison of evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16(4)* 555-568.

Fisher, C. (1996). Structural limits on verb mapping: the role of analogy in children's interpretations of sentences. *Cognitive Psychology, 31* 41-81.

Fisher, C. (2002,a). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition*, 82, 259-278.

Fisher, C. (2002,b). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science, 5* 56–65.

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology 23*, 331-392.

Frazier, L., & Fodor, J. D. (1978). The Sausage Machine: A new two-stage parsing model. *Cognition, 6* 291-325.

Frazier, L. (1987). Sentence processing: a tutorial review. In M. Colthear (Ed.), *Attention and Performance XII: The psychology of reading* (pp. 559-586), Hillsdale NJ: Erlbaum.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14* 178-210.

Garnsey, S. M., Pearlmutter, N. J., Meyers, E., & Lotocky, M. A. (1997). The contribution of verb bias to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37* 58-93.

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

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17,* 648-691.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135-176.

Givon, T. (1985). Function, structure and language acquisition. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (Vol. 2, pp. 1005-1028), Hillsdale, N.J.: Lawrence Elbaum Associates.

Gleitman, L. (1994). The structural sources of verb meanings. In P. Bloom (Ed.), *Language acquisition: Core readings.* Cambridge, MA: MIT Press.

Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*, 23-64.

Gleitman, L. R. (1990). The structural sources of word learning. *Language Acquisition, 1* 3-55.

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

Goldberg, A. E. (2004). But do we need universal grammar? Comment on Lidz et al. (2003). *Cognition, 94*, 77-84.

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

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature, 391*, 279-281.

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

Green, G.M. (1974). *Semantics and syntactic regularity.* Bloomington: Indiana University Press.

Grimshaw, J. (1981). Form, function, and the language acquisition device. In C. L. Baker & J. McCarthy (Eds.) *The logical problem of language acquisition.* (pp. 165-182.), Cambridge, 1981: MIT Press.

Grimshaw, Jane. 1990. *Argument structure.* Cambridge, Mass.: MIT Press.

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.

Gropen, J., Pinker, S, Hollander, J., & Goldberg, R. (1991a). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language, 18*, 115-151.

Gropen, J., Pinker, S, Hollander, J., & Goldberg, R. (1991b). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition, 41*, 153-195.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp.10-1 -10-112). New York: Wiley.

Hare, M., McRae, K., & Elman, J. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language, 48*, 281-303.

Hare, M., McRae, K., & Elman, J. L. (2004). Admitting that admitting verb sense into the corpus analyses makes sense. *Language and Cognitive Processes, 19*, 181-224.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization.* Cambridge, England: Cambridge University Press.

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.

Jackendoff. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

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. (1993). Contingent frequency effects in syntactic ambiguity resolution. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society,* 593-598. Erlbaum, Hillsdale, NJ.

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

YOU NEED TO ADD KEMP, PERFORS AND TENENBAUM

Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review,* 8, 132–138

Kim, J., Bangalore, S., & Trueswell, J.C. (2002). The convergence of lexicalist perspectives in psycholinguistics and computational linguistics. In P. Merlo & S. Stevenson (Eds.), *Lexical basis for sentence processing*. Philadelphia: John Benjamins.

Koenig, J.-P., Mauner, G., & Bienvenue, B. (2003). Arguments for adjuncts. *Cognition, 89*, 67-103.

Leiven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: a usage based approach. *Journal of Child Language, 30*, 333-370.

Leiven, E., Pine, J., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*, 187-220.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition, 87*, 151-178.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Magnuson, J. S., Tanenhaus, M. K. and Aslin, R. N. (2002). Immediate integration of syntactic and referential constraints on spoken word recognition. In W. D. Gray & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, 614 - 619. Mahway, NJ: Erlbaum

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*, 202-227.

Marcus, G., Vijayan, S., Rao, S., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77-80.

Mintz, T. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition, 30*, 678- 686.

Mintz, T., Newport, E., & Bever, T. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393-424.

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*. Hillsdale, NJ: Erlbaum, 601-681

Moeser, D. S., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior, 11*, 759-769.

Morgan, J., Meier, R., & Newport, E. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498-550.

Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20*, 67-85.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17,* 357-374.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Snow and C. A. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition*., Cambridge: Cambridge University Press.

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

Pinker, S. (1984). *Language learnability and language development*. Cambridge, Mass: Harvard University Press.

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

Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26*, 195-267.

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. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

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

Singleton, J.L., & Newport, E.L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology, 49,* 370-407.

Snedeker, J. (2000). Cross-situational observation and the semantic bootstrapping hypothesis. *Proceedings of the Thirtieth Annual Child Language Research Forum,* Stanford, C.A.: Center for the Study of Language and Information.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49,* 238-299.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes, 12,* 211-272.

Tanenhaus, M., & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of Perception and Cognition* (Vol. 11). New York: Academic Press.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629-641.

Tenenbaum, J. B., Griffiths, T. L., and Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10,* 309-318.

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

Thompson, S.P., & Newport, E.L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3,* 1-42.

Tomasello, M. (1992). *First verbs: A case study in early grammatical development*. Cambridge: Cambridge University Press.

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

Tomasello, M. & Abbot-Smith, K. (2002). A tale of two theories: Response to Fisher. *Cognition*, 83, 207-214.

Tomasello, M., Akhtar, N., Dodson, K., & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language, 24,* 373-387.

Trueswell, J. C., & Kim, A. E. (1998). How to prune a garden-path by nipping it in the bud: Fast-priming of verb argument structures. *Journal of Memory and Language, 39,* 102-123.

Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying online sentence processing in young children. *Cognition,73,* 89-134.

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

Author's Note

This research was supported by National Institute of Health research grants DC-00167 and NIH HD-27206 awarded to Elissa Newport and Michael Tanenhaus respectively, and by ESRC grant PTA-026-1296 awarded to Elizabeth Wonnacott. Many thanks to Edward Longhurst, who wrote the ExBuilder software which ran these experiments, to Dr. Ted Supalla and Don Metlay for providing their time and resources to create the video stimuli, to those working in the lab who were involved in creating stimuli and running participants: Dana Subik, Whitney Hopfinger, Carol Faden, Maggie Chang, Catherine Craft, Katie Schuler, Joyce Akwaa and Katie Dickerson, and to Dr. Jeff Runner and Dr. Joyce McDonough for their insightful comments on these topics.

Footnotes

(1) Acquiring these construction meanings also present a non-trivial problem – an issue which has recently been addressed in work by Goldberg and colleagues (Casenhiser & Goldberg, 2005; Goldberg, Casenhiser, & Sethuraman, 2004). In a similar vein to our own work, these experiments taught adult and child learners novel verbs within the context of a new construction (SOV) with a new meaning ('subject appears on object via action denoted by the verb'). The work demonstrates that learning the meaning of a construction is facilitated when there is a disproportionately high frequency of occurrence of that construction with one particular verb, thus demonstrating an ability to perform simple distributional analyses of the type also explored in this work. In the experiments we report here the two constructions have the same meaning which is deliberately generic to minimize this aspect of learning.

(2) Participants occasionally mispronounced one of the two names. However, in every case it was possible to identify at least one of the nouns, so that we were able to successfully identify the word order. Word orders were coded and scored by a coder who had no knowledge of verb class.

(3) Unpublished experiments involving the constructions *VAP* and *VPA* (i.e. without the *ka*) have found the same *VAP* bias, suggesting that it is not the presence of the particle which makes the *VPA_ka* construction dispreferred.

(4) In this way these test items differed from the 'new' test items used in the artificial language learning experiment presented in Wonnacott and Newport (2005), which

contained nouns only heard in vocabulary training. That choice resulted in quite different findings, possibly because the usage of an entirely new vocabulary discouraged participants from focusing on distributional information.

(5) In the lexical language, one participant used the incorrect V**AP**_ka construction consistently with one particular verb. Since this behavior was systematic, we classified it as 'correct' and counted these productions as though they were VAP constructions. Statistics were performed both with these data scored as VAP productions (the figures reported the text) and with these scores excluded from the analyses, yielding an identical pattern of results.

(6) As with novel verbs, in the *lexical* language one participant used the incorrect *VAP_ka* construction consistently with one particular verb, and again we have counted these as *VAP* productions.

(7) We recognize that our learners did not entirely avoid over-generalization with one-construction verbs. However, this is only problematic if we assume that the end product of syntax acquisition is a grammar that unequivocally rejects ungrammatical forms. Recent work suggests that even adult native speakers may be graded in their linguistic judgments, particularly with low frequency verbs (e.g. Theakston, 2004). Thus the fact that our learners, who have received only five days of exposure to these languages, show over-generalization, is consistent with natural language acquisition. The frequency effects suggest that learning is following a normal developmental path, and we expect that increased exposure to our languages would make learners increasingly less likely to over-generalize.

(8) We thank an anonymous reviewer for this suggestion.

Table 1: Experiment 1 Grammaticality Judgment Test: Proportion of sentence types accepted as grammatical for each verb type.

		Sentence Type		
	Verb Class	***VAP***	***VPA_ka***	***Difference***
***VAP*-only**	**low frequency**	98.2%	67.9%	*30.4%*
***VAP*-only**	**High frequency**	100.0%	39.4%	*60.7%*
***VPA_ka*-only**	**low frequency**	53.6%	96.4%	*-42.9%*
***VPA_ka*-only**	**High frequency**	46.4%	100.0%	*-53.6%*
alternating	**medium frequency**	88.4%	94.6%	*-6.3*

Table 2: Experiment 2 Grammaticality Judgment Test: Proportion of sentence types accepted as grammatical for each verb type.

		Sentence Type		
Verb Class		***VAP***	***VPA_ka***	***Difference***
***VAP*-only**	**low frequency**	100.00%	81.6%	18.4%
***VAP*-only**	**high frequency**	96.2%	81.3%	14.8%
VPA_ka only	**low frequency**	77.8%	96.4%	-18.7%
VPA_ka only	**high frequency**	74.2%	96.4%	-22.3%
alternating	**across frequencies**	97.3%	93.3%	4.0%

Table 3: Experiment 3 Grammaticality Judgment Test: Proportion of sentence types accepted as grammatical for each verb type in each language condition.

	Sentence Type		
Verb Class	***VAP***	***VPA_ka***	**Difference**
***lexical* language: *VAP*-only**	100.0%	23.3%	-76.7%
***lexical* language: *VAP*-only**	17.9%	98.3%	80.4%
***lexical* language: novel verbs**	72.5%	100.0%	27.5%
lexical* language: minimal exposure *VAP	100.0%	5%	-95.0%
lexical* language: minimal exposure *VPA_ka	10.0%	90%	80.0%
***generalist* language: alternating biased**	80.7%	98.0%	17.3%
***generalist* language: novel verbs**	62.5%	60.0%	-2.5%
generalist* language: minimal exposure *VAP	90.0%	90.0%	0.0%
generalist* language: minimal exposure *VPA_ka	90.0%	90,0%	0.0%

Figure Captions

Figure 1: Screen shot from Online Comprehension Test. At this point in the trial, the two videos have each played in turn and now remain on the screen, frozen in the final frame. The participant will now use the mouse to click on the white cross between the two.

Figure 2: Experiment 1 Production Test: Proportion of correct productions with *VAP* construction for each verb type.

Figure 3: Experiment 1 Grammaticality Judgment Test: Difference scores showing preference for *VAP* construction.

Figure 4: Experiment 1 Online Comprehension Test: Pattern of looks for (from top left to bottom right): *VPA_ka* sentence with *VPA_ka*-only verb, *VAP* sentence with *VAP*-only verb, *VPA_ka* sentence with alternating verb, *VAP* sentence with alternating verb, (boundaries allow 200ms delay for initiating an eye-movement).

Figure 5: Experiment 1 Online Comprehension: Percentage of looks to *VAP* picture during critical region of sentence (i.e. after beginning of first noun and before 'ka').

Figure 6: Experiment 2 Production Test: Proportion of correct productions with *VAP* construction for each verb type.

Figure 7: Experiment 2 Grammaticality Judgment Test: Difference scores showing preference for *VAP* construction.

Figure 8: Experiment 2 On-line Comprehension Test: Percentage of looks to *VAP* picture during critical region of sentence (from beginning of first noun until 'ka').

Figure 9: Experiment 3 Production Test: Proportion of productions with *VPA_ka* construction for each verb type in the *lexical* condition (above) and *generalist* condition (below).

Figure 10: Experiment 3 Grammaticality Judgment Test: Difference scores showing preference for *VPA_ka* construction for each verb type in the *lexical* condition (above) and *generalist* condition (below).

Figure 11: Experiment 3 Online Comprehension Test: Pattern of looks for (from top left to bottom right): *Lexical* Condition *VPA_ka* sentence with *VPA_ka*-only verb, *Lexical* Condition *VAP* sentence with *VAP*-only verb, *Generalist* Condition *VPA_ka* sentence with biased alternating verb, *Generalist* Condition *VAP* sentence with biased alternating verb (boundaries allow 200ms delay for initiating an eye-movement).

Figure 12: Percentage of looks to *VPA_ka* construction during critical region of sentence (from beginning of first noun until 'ka') for familiar verbs in the *Lexical* condition (above) and *Generalist* condition (below).

Figure 13: Experiment 3 Online Comprehension Test: Pattern of looks for (from top left to bottom right): *Lexical* Condition *VPA_ka* sentence with minimal exposure *VPA_ka* verb, *Lexical* Condition *VAP* sentence with minimal exposure *VAP* verb, *Generalist* Condition *VPA_ka* sentence with minimal exposure *VPA_ka* verb, *Generalist* Condition

VAP sentence with minimal exposure *VAP* verb, (boundaries allow 200ms delay for initiating an eye-movement).

Figure 1

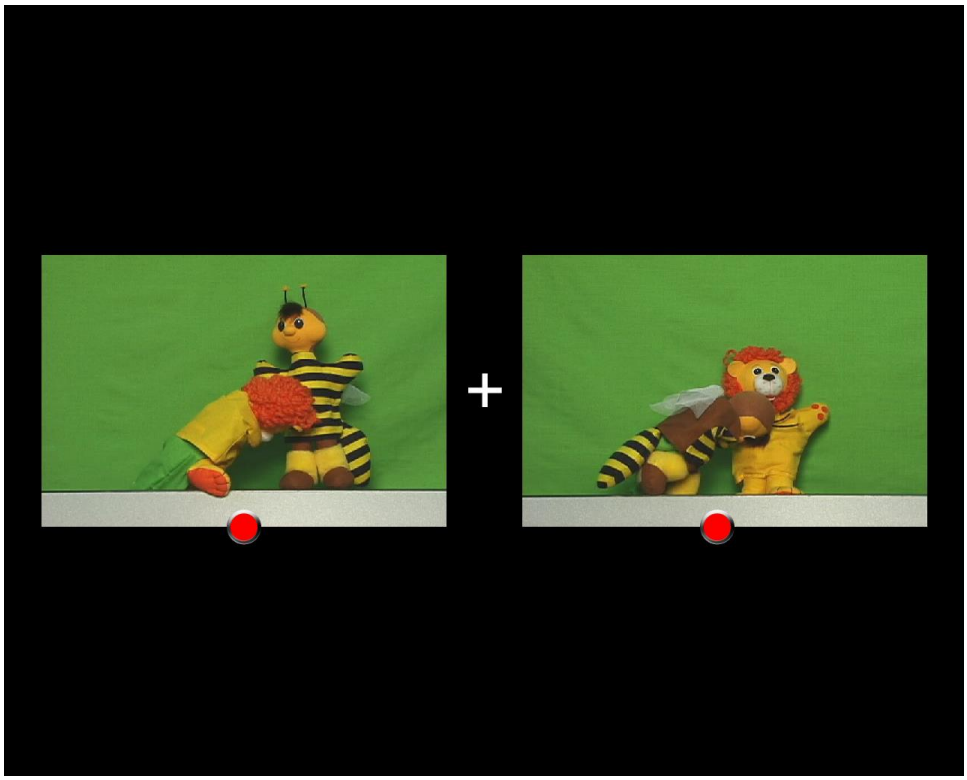


Figure 2

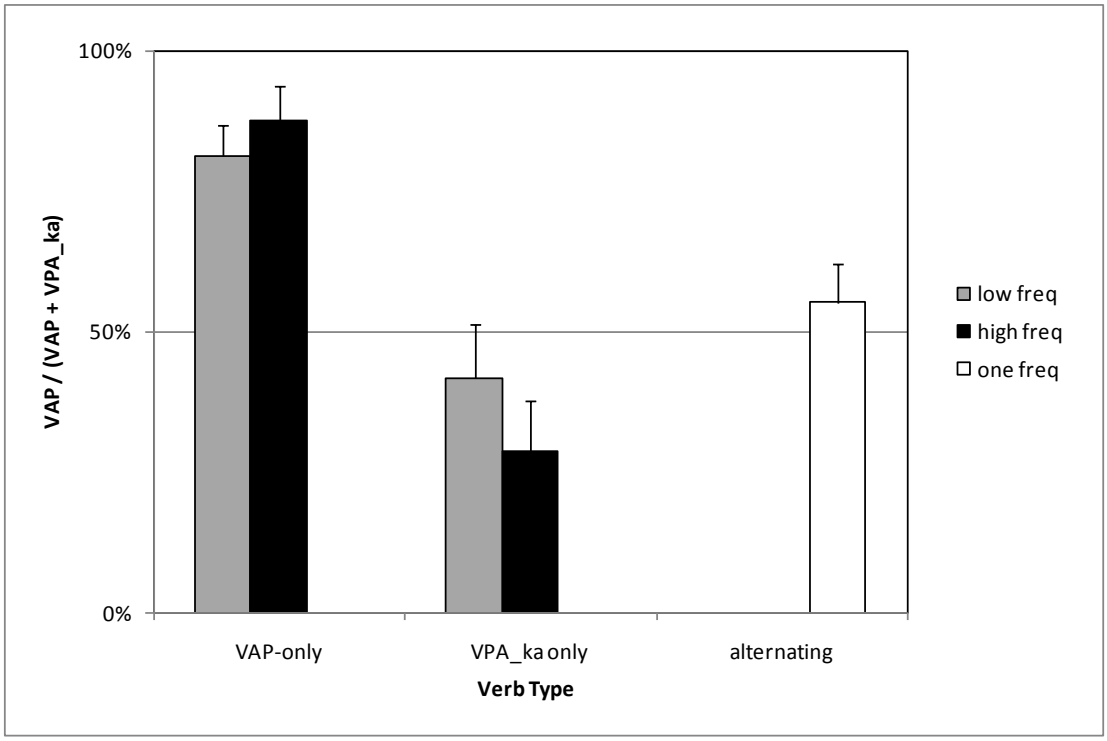

100%
50%
0%
VAP / (VAP + VPA_ka)
VAP-only
VPA_ka only
alternating
Verb Type
low freq
high freq
one freq

Figure 3

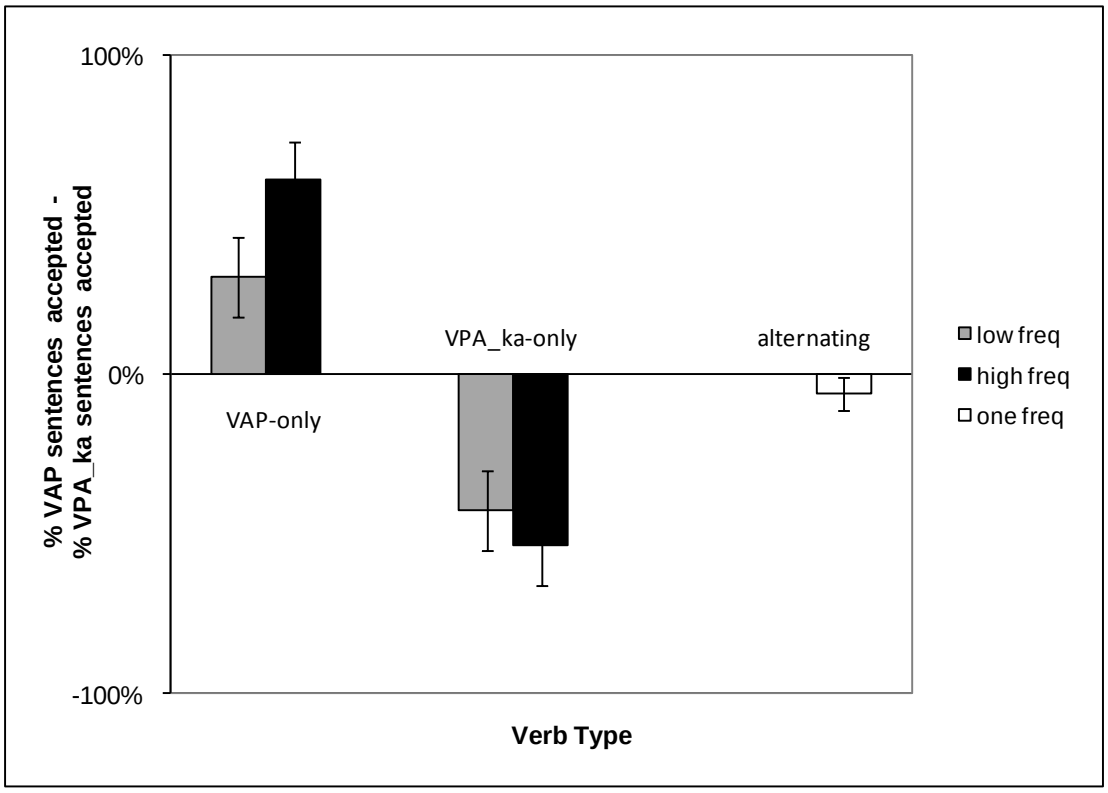

100%
0%
-100%
% VAP sentences accepted - % VPA_ka sentences accepted
VAP-only
VPA_ka-only
alternating
low freq
high freq
one freq
Verb Type

Figure 4

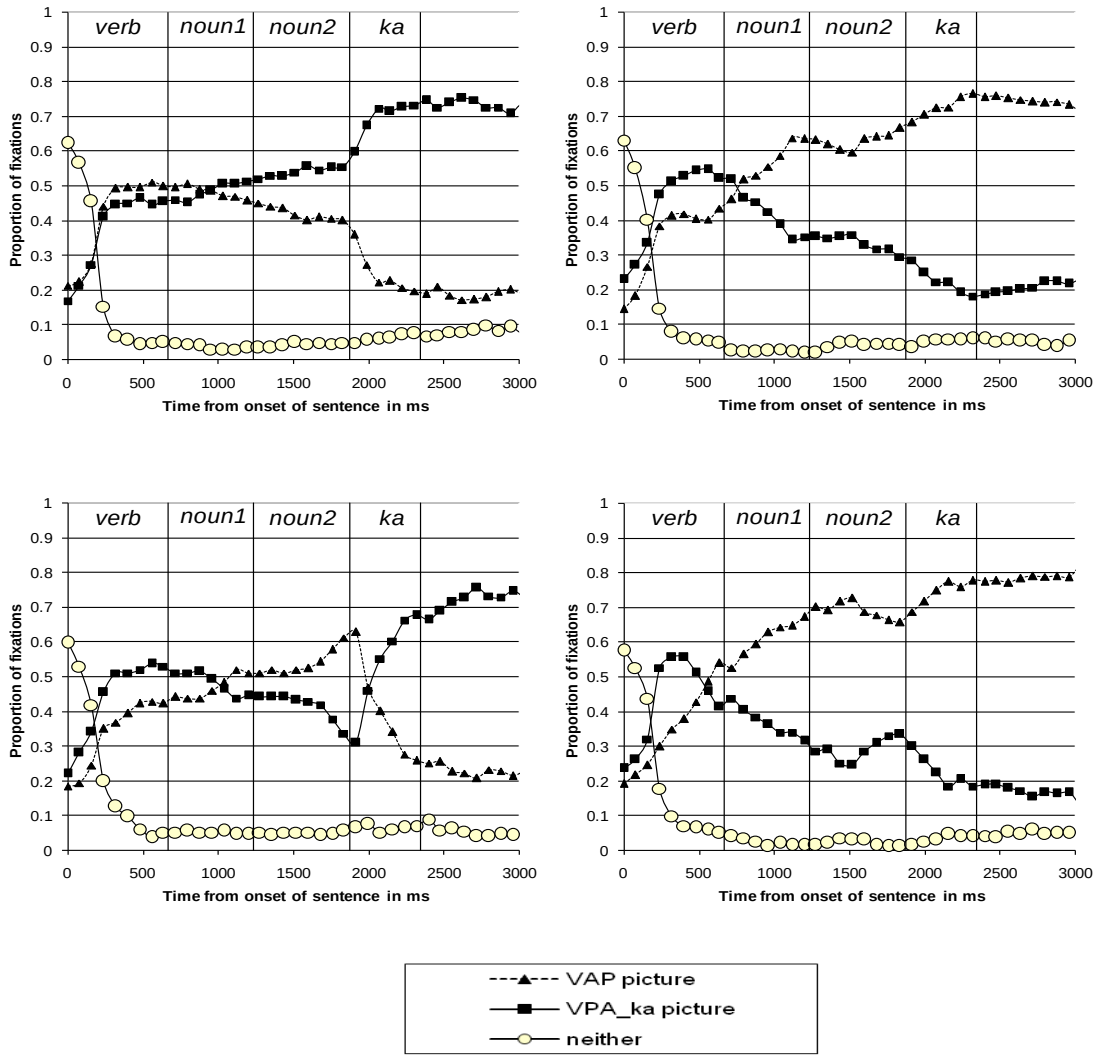

verb
noun1
noun2
ka
Proportion of fixations
Time from onset of sentence in ms
VAP picture
VPA_ka picture
neither

Figure 5

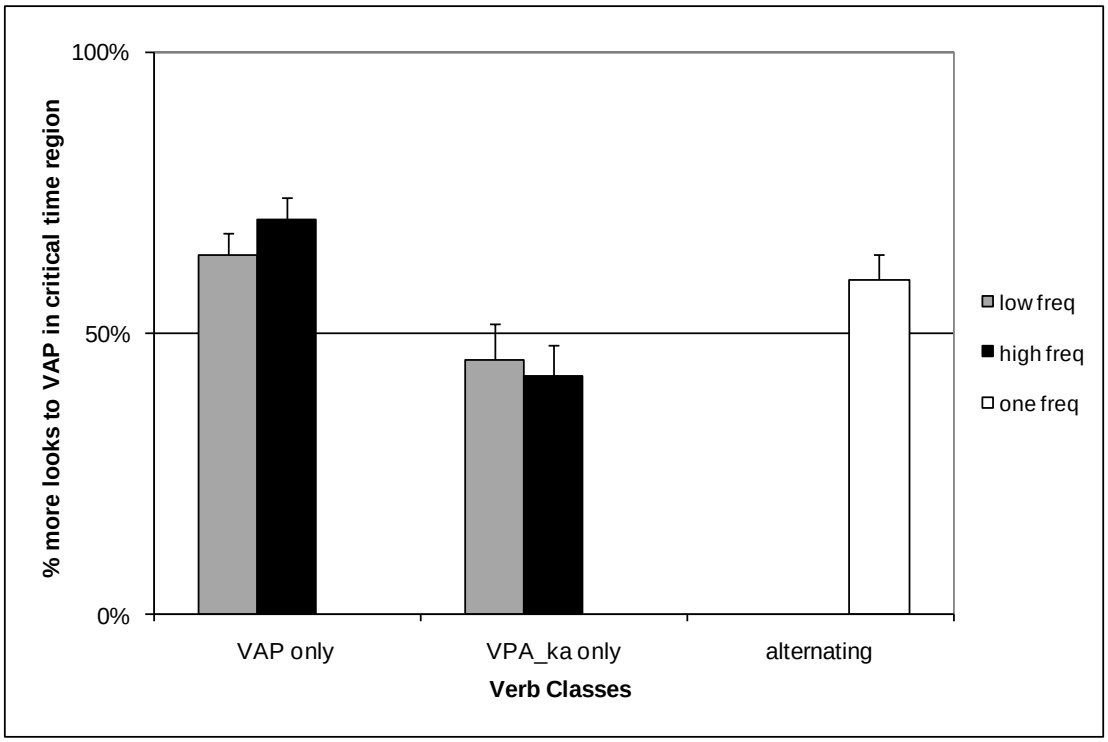

100%
50%
0%
% more looks to VAP in critical time region
VAP only
VPA_ka only
alternating
Verb Classes
low freq
high freq
one freq

Figure 6

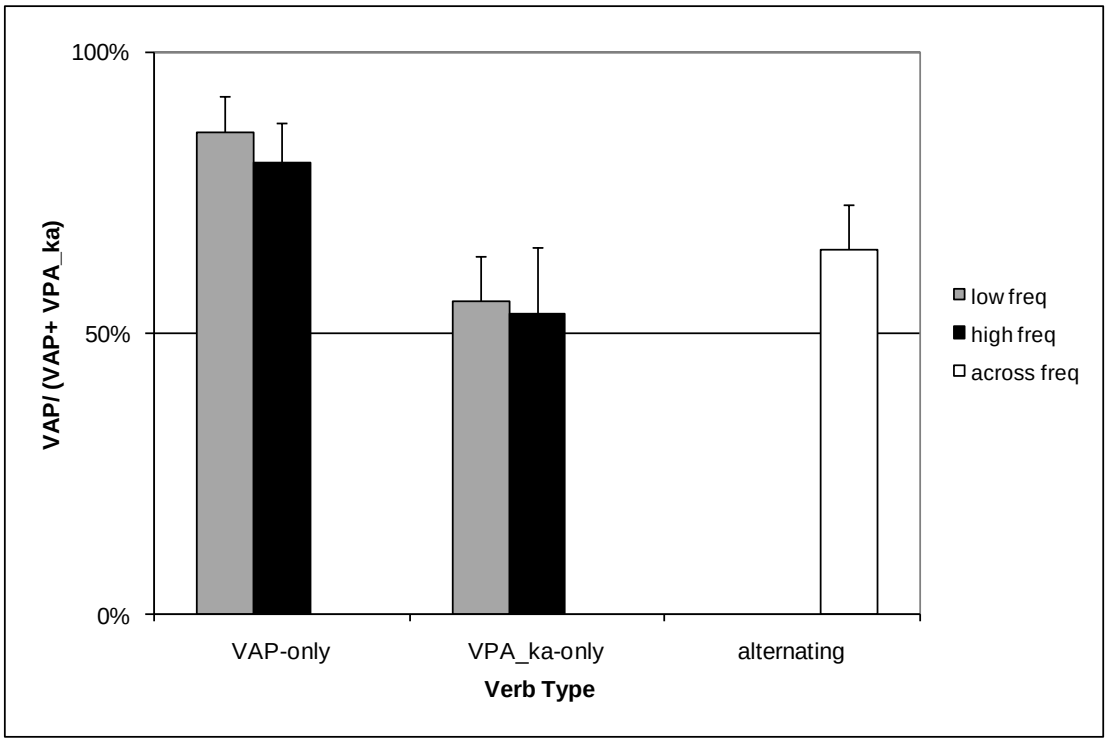

100%
50%
0%
VAP/ (VAP+ VPA_ka)
VAP-only
VPA_ka-only
alternating
Verb Type
low freq
high freq
across freq

Figure 7

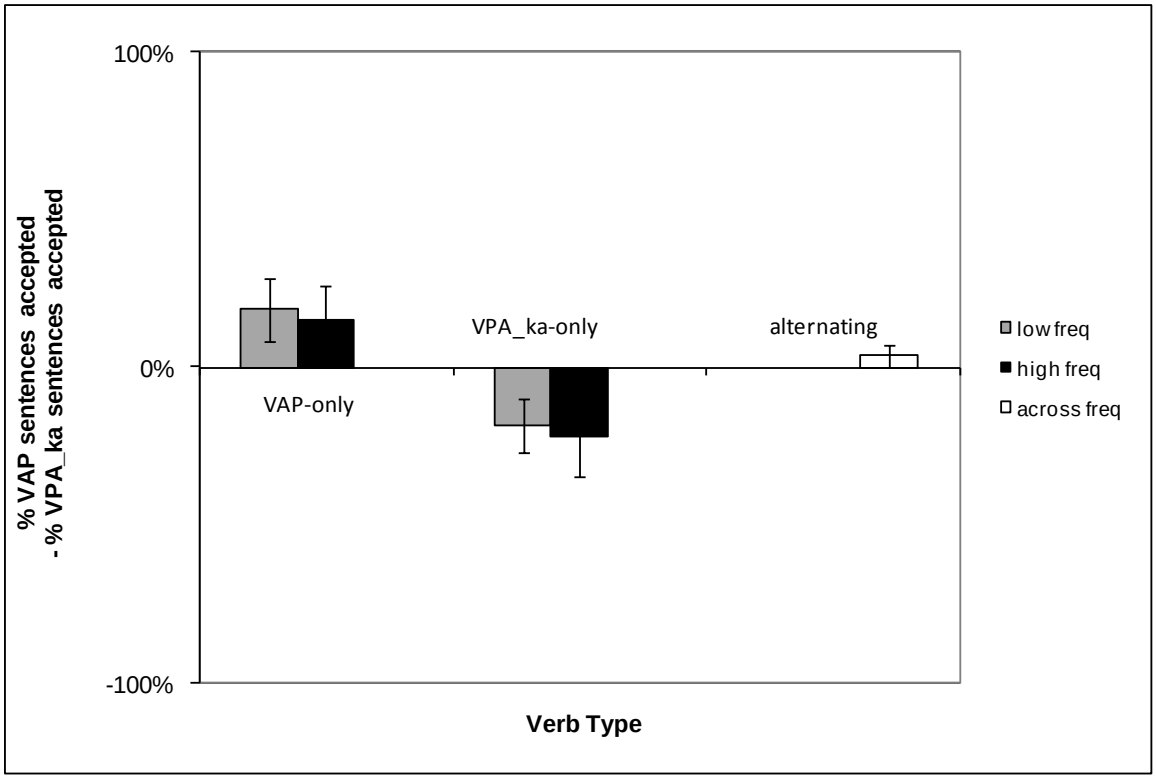

100%
0%
-100%
% VAP sentences accepted
- % VPA_ka sentences accepted
VAP-only
VPA_ka-only
alternating
low freq
high freq
across freq
Verb Type

Figure 8

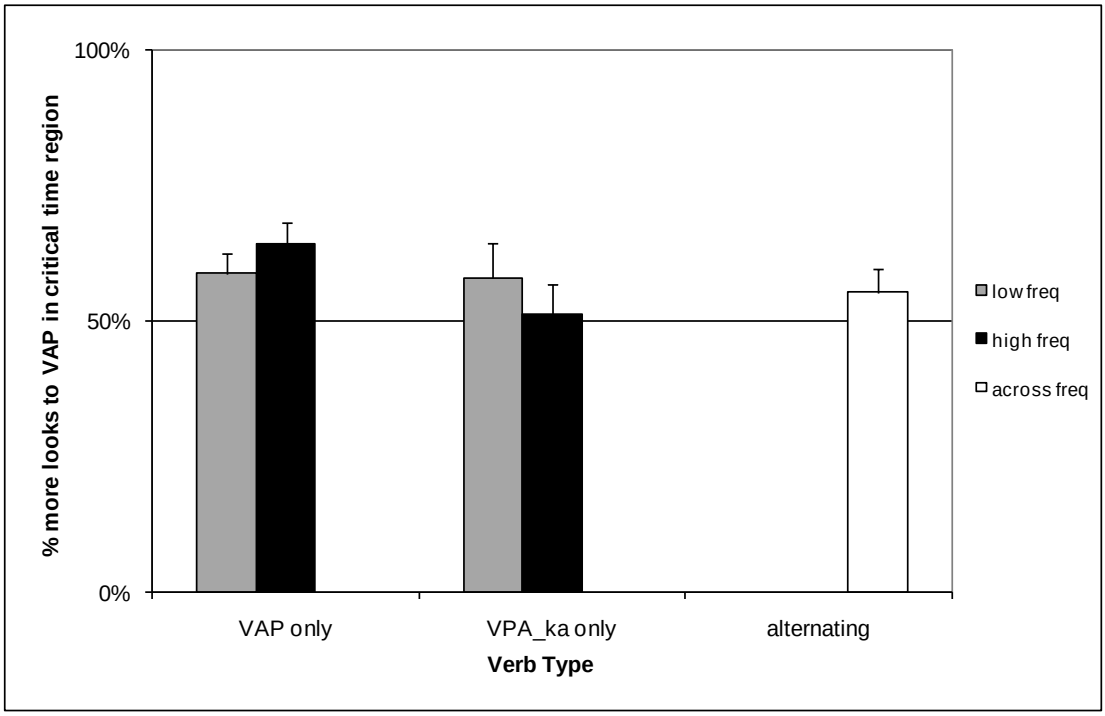

100%
50%
0%
% more looks to VAP in critical time region
VAP only
VPA_ka only
alternating
Verb Type
low freq
high freq
across freq

Figure 9

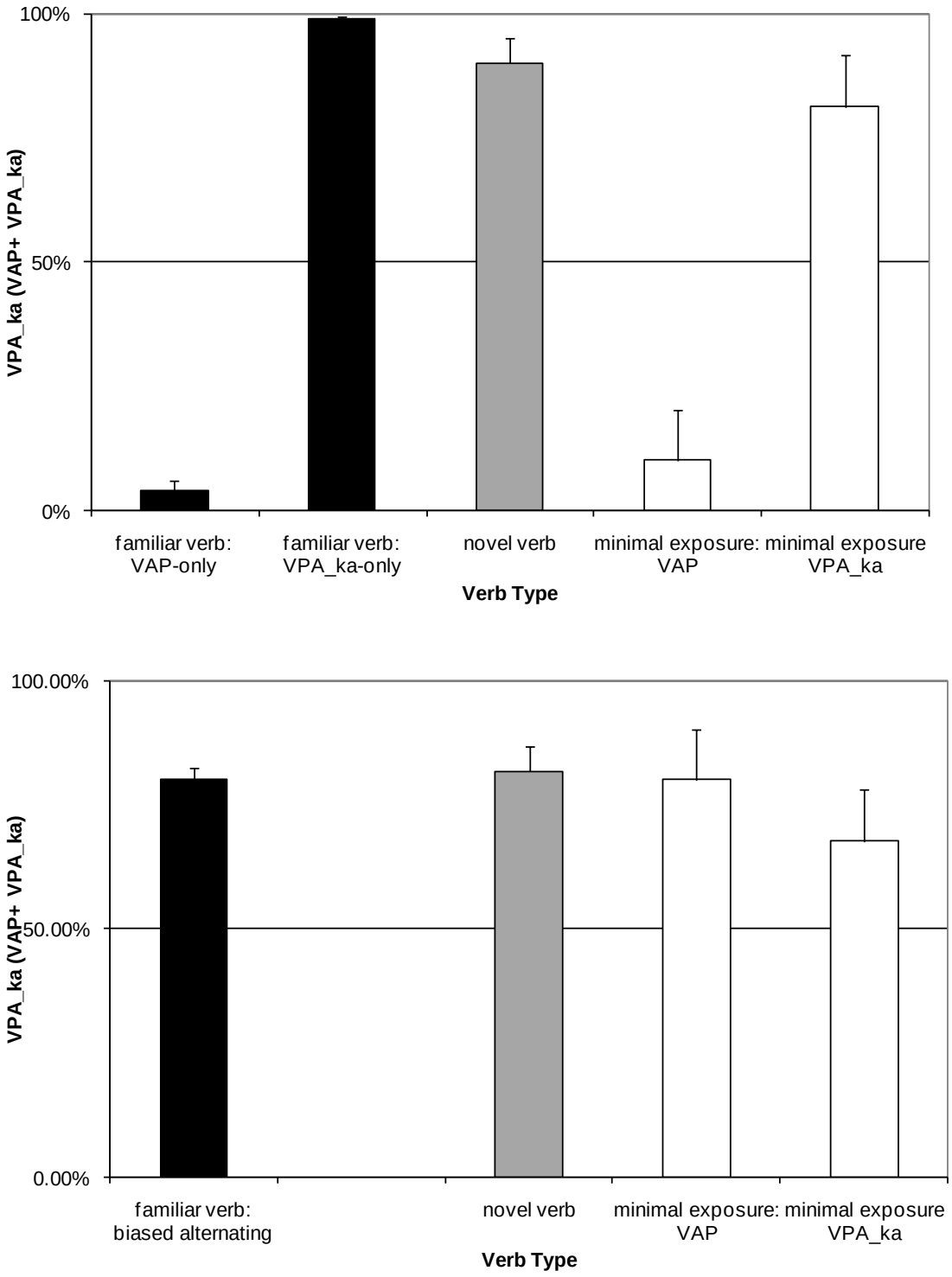
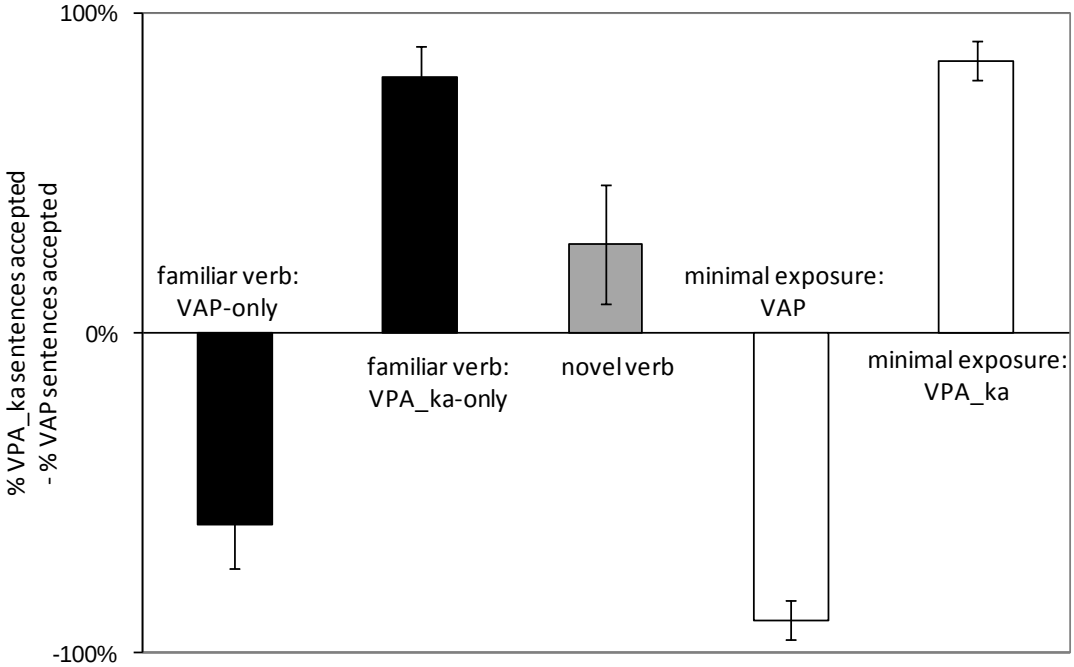

100%
50%
0%
VPA_ka (VAP+ VPA_ka)
familiar verb: VAP-only
familiar verb: VPA_ka-only
novel verb
minimal exposure: VAP
minimal exposure: VPA_ka
Verb Type
100.00%
50.00%
0.00%
VPA_ka (VAP+ VPA_ka)
familiar verb: biased alternating
novel verb
minimal exposure: VAP
minimal exposure: VPA_ka
Verb Type

Figure 10


100%
0%
-100%
% VPA_ka sentences accepted - % VAP sentences accepted
familiar verb: VAP-only
familiar verb: VPA_ka-only
novel verb
minimal exposure: VAP
minimal exposure: VPA_ka

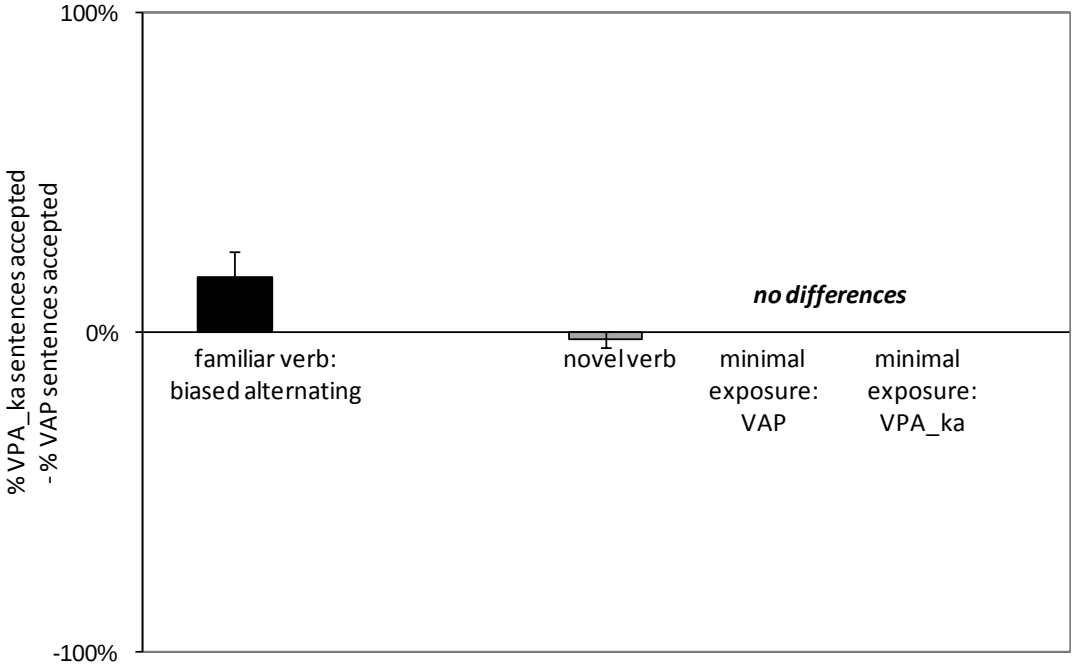

100%
0%
-100%
% VPA_ka sentences accepted - % VAP sentences accepted
no differences
familiar verb: biased alternating
novel verb
minimal exposure: VAP
minimal exposure: VPA_ka

Figure 11

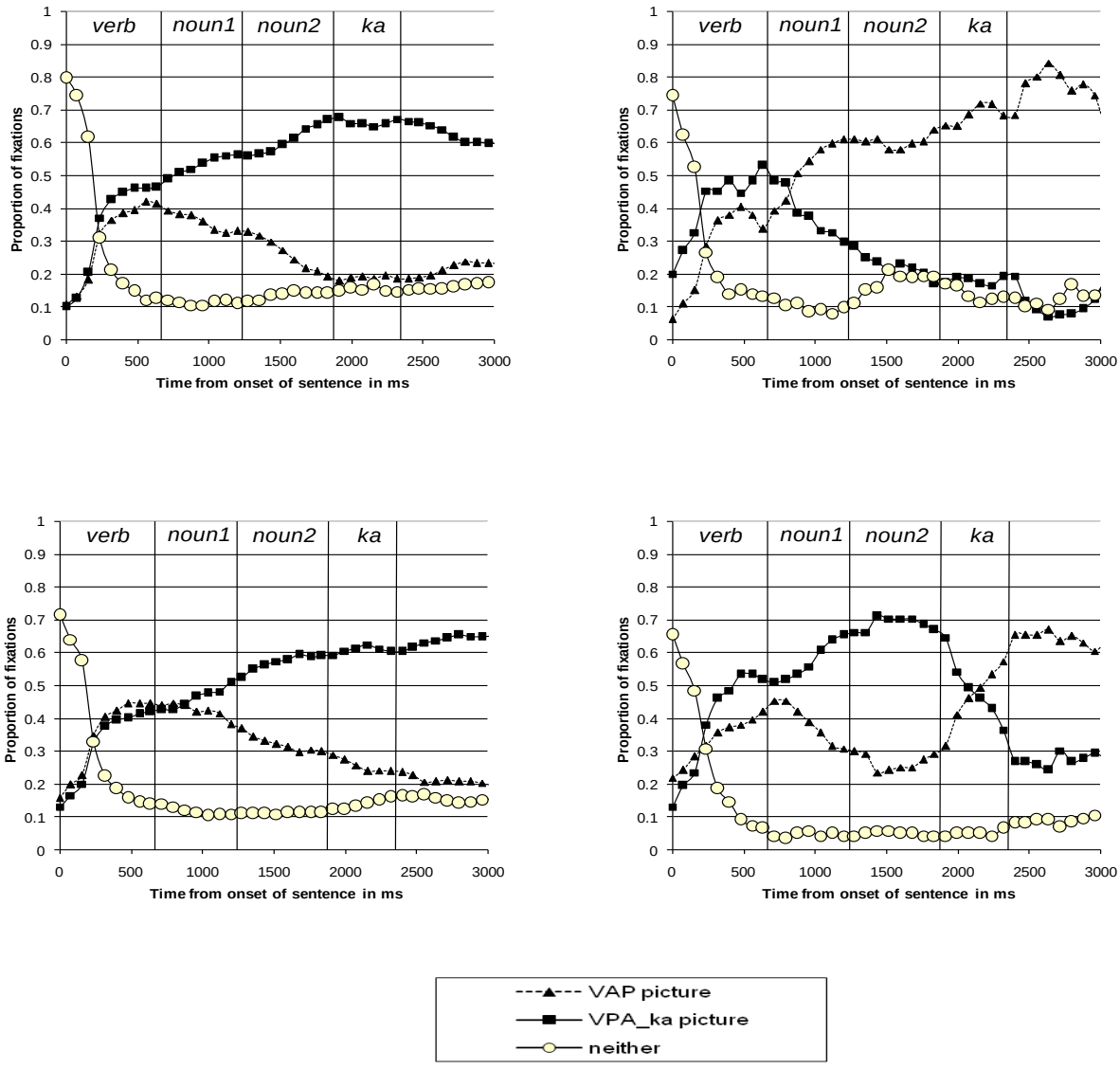
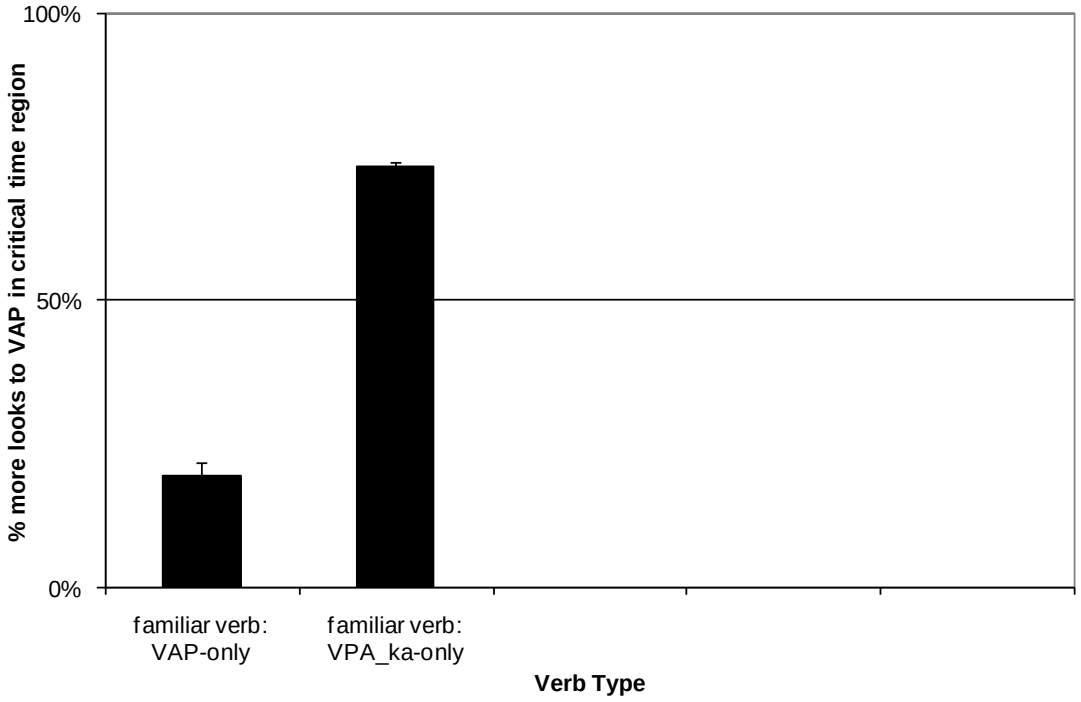

verb
noun1
noun2
ka
Proportion of fixations
Time from onset of sentence in ms
VAP picture
VPA_ka picture
neither

Figure 12


100%
50%
0%
% more looks to VAP in critical time region
familiar verb: VAP-only
familiar verb: VPA_ka-only
Verb Type

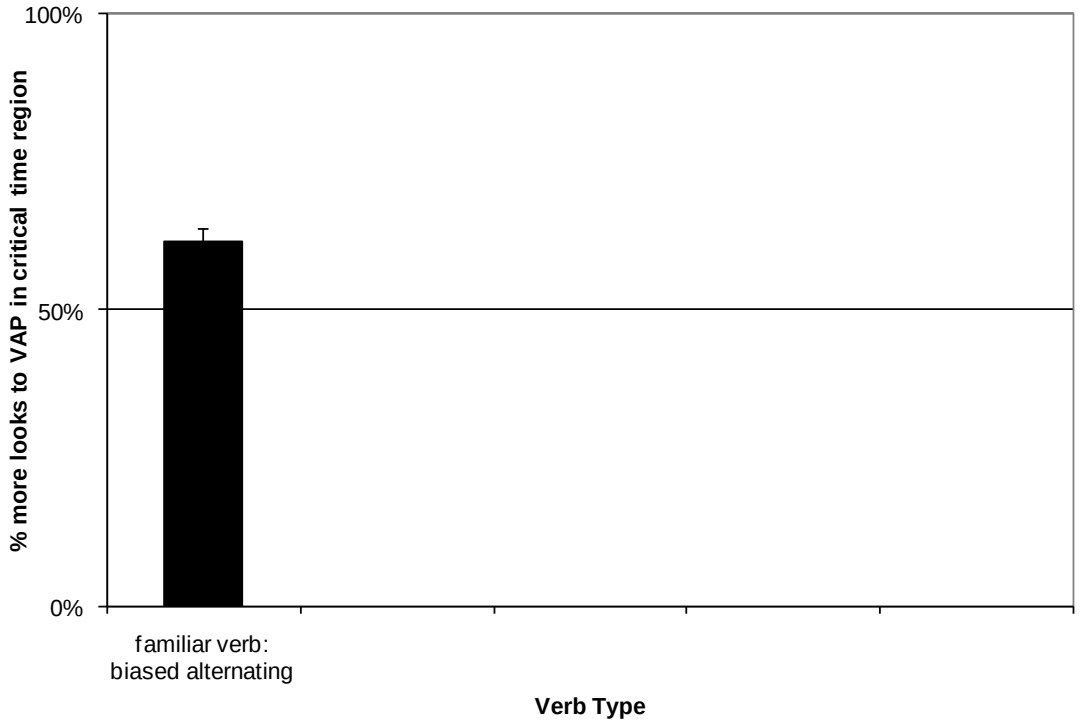
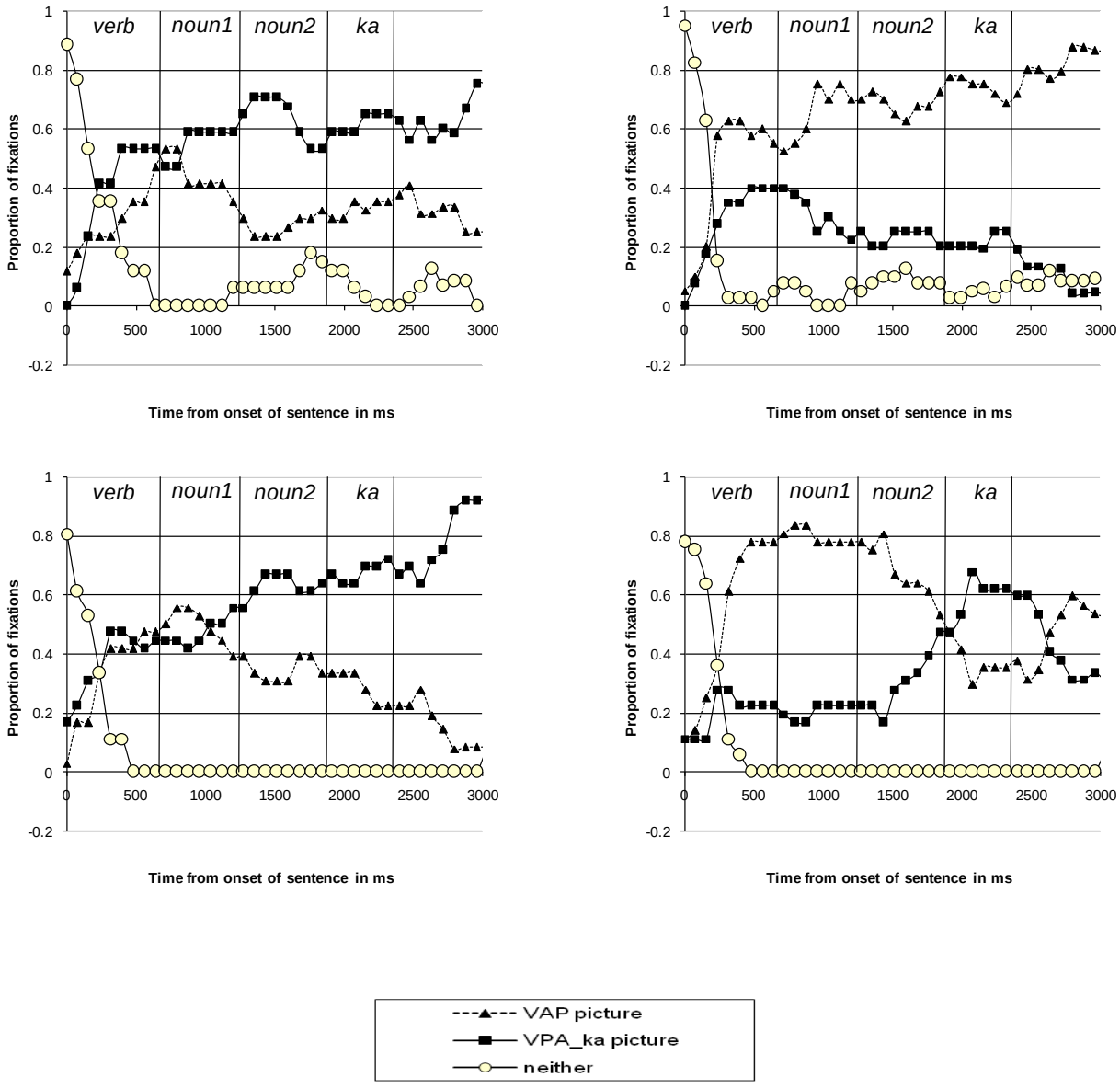

100%
50%
0%
% more looks to VAP in critical time region
familiar verb: biased alternating
Verb Type

Figure 13


verb
noun1
noun2
ka
Proportion of fixations
Time from onset of sentence in ms
VAP picture
VPA_ka picture
neither