ARTICLE IN PRESS

Journal of Memory and Language xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

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

Input effects on the acquisition of a novel phrasal construction in 5 year olds

Elizabeth Wonnacott [a,*], Jeremy K. Boyd [b], Jennifer Thomson [a], Adele E. Goldberg [c]

[a] Department of Experimental Psychology, South Parks Road, Oxford, OX1 3UD, United Kingdom
[b] Center for Research in Language, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA 92093-0526, United States
[c] Department Of Linguistics, Princeton University, Princeton, NJ 08544, United States

ARTICLE INFO

Article history:
Received 3 July 2010
revision received 9 November 2011
Available online xxxx

Keywords:
Verb argument structures
Artificial language learning
Novel construction learning
Statistical learning

ABSTRACT

The present experiments demonstrate that children as young as five years old ($M = 5{:}2$) generalize beyond their input on the basis of minimal exposure to a novel argument structure construction. The novel construction that was used involved a non-English phrasal pattern: VN_1N_2, paired with a novel abstract meaning: N_2 approaches N_1. At the same time, we find that children are keenly sensitive to the input: they show knowledge of the construction after a single day of exposure but this grows stronger after 3 days; also, children generalize more readily to new verbs when the input contains more than one verb.

Introduction

A characteristic property of natural languages is the systematic correlation between structural patterns and abstract semantic or information structure functions (Fillmore, 1968; Grimshaw, 1990; Landau & Gleitman, 1985; Pinker, 1989). Such correspondences in the domain of argument structure—encapsulated by the notion of argument structure *constructions*—provide the basic clause types of a language (Goldberg, 1995). For example, the English sentences *Katie gave Jack the book* and *Poppy baked Henry a cake* are both instances of the ditransitive construction—a common phrasal pattern involving a subject and two objects. The two sentences contain distinct words but both convey actual or intended transfer. Our knowledge of this abstract linking is evident in the fact that we can use the construction productively—i.e., it can be used with new lexical items that may or may not lexically encode the transfer meaning. For example, if asked what *She mooped him something* means, speakers are quite likely to guess that she *gave* him something (Ahrens, 1995; Goldberg, 1995). In fact, adults generally interpret utterances with novel verbs by attending to the semantics of the argument structure constructions involved (Goldwater & Markman, 2009; Johnson & Goldberg, submitted for publication a; Kako, 2006; Kaschak & Glenberg, 2000).

At the same time, there is a question about whether young children are able to use argument structure constructions in the same way as adults. There is a great deal of evidence that children's early productions tend to avoid straying too far from their input. For example, when children younger than three hear a novel verb used intransitively, they are highly unlikely to productively transitivize it (Akhtar & Tomasello, 1997; Baker, 1979; Bates & MacWhinney, 1987; Braine, 1976; Pinker, 1989; Tomasello, 2000). Such experimental data, along with data from spontaneous production (Bowerman, 1982; Ingram & Thompson, 1996; Lieven, Pine, & Baldwin, 1997; Tomasello, 1992), have led to the proposal that early grammars lack abstract argument structure representations and that apparent uses of a construction actually rely on verb-specific representations (so called *verb-islands*; Tomasello, 2000).

Evidence from comprehension is somewhat more mixed. Experiments using the act out procedure, in which

* Corresponding author. Address: Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom.
E-mail address: E.A.Wonnacott@warwick.ac.uk (E. Wonnacott).

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.11.004

the child is required to demonstrate knowledge of a construction by acting it out with puppets, have again found that children younger than three have difficulty extending a new verb from an intransitive to a transitive construction (e.g., Akhtar & Tomasello, 1997). On the other hand, experiments using the preferential looking paradigm have demonstrated that young children can use knowledge of the semantics of a frame to make inferences about the meaning of a new verb (Fisher, 1996, 2002; Naigles, 1990; see Gleitman, Cassidy, Nappa, Papafragou, and Trueswell (2005) for a review). There is also evidence that children as young as 21-months have some knowledge of the link between specific word order and the abstract semantics of the English transitive construction (Gertner, Fisher, & Eisengart, 2006). Other work has found that young children require scaffolding in the form of initial exposure to familiar verbs used transitively in order to demonstrate any knowledge of the generalization (Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008); this finding indicates that early generalizations may initially be tentative or "graded" (Abbot-Smith, Lieven, & Tomasello, 2008).

Given the evidence that lexically specific constructions are characteristic of young children's early productions and, at least to some extent, their early comprehension as well, it is important to ask what sorts of input ultimately encourage generalization. Novel construction learning studies allow us to manipulate the input systematically, so that we can explore how the structure of that input affects the nature of the abstractions acquired. Wonnacott, Newport, and Tanenhaus (2008) explored the effect of input structure on the generalization of two novel constructions to unattested verbs, using an artificial language learning paradigm (cf. also e.g., Braine, 1963; Gomez, 2002; Hudson Kam & Newport, 2005, 2009). In the context of learning a novel artificial language, over five learning sessions, adult learners were exposed to a set of sentences exemplifying two novel phrasal patterns (VN_1N_2 and VN_2N_1 *particle*) which each mapped to a semantic pattern where the entity denoted by N_1 was the *agent* of an action which affected the entity denoted by N_2 (the *patient*). Note that this is the same basic semantics denoted by common examples of the familiar English transitive construction. Using a variety of methods, learners were tested on their usage and acceptance of the two constructions with attested and unattested verbs. The results demonstrated strong learning of both constructions, and a clear ability to extend the constructions to unfamiliar verbs (i.e., novel verbs not used in the exposure sentences). However, the statistical structure of the input affected the extension of familiar verbs (used in the exposure sentences) from one construction to the alternative construction. For example, learners were less likely to extend verbs that had frequently appeared in the alternative construction, and were also less likely to extend constructions given a language in which there was strong evidence that the usage of constructions was lexically conditioned. Wonnacott et al. argued that the generalization of constructions to new verbs depended upon the input in a rational, evidence-based manner, and Perfors, Tenenbaum, and Wonnacott (2010) demonstrated that human performance is in line with the predictions of a hierarchical Bayesian model. This pattern of learning has also been demonstrated in children, though in a different linguistic domain not involving verb argument structures (Wonnacott, 2011), suggesting that this type of learning may be relevant for language acquisition.

The results of Wonnacott et al. (2008) suggest that generalization is a function of the statistical structure of the input. However one limitation of the study from the perspective of exploring novel construction learning per se, is that it does not consider the situation in which a novel phrasal form is associated with a novel abstract meaning (i.e., one not encoded by any existing English construction). Certain previous studies that have used a familiar meaning encoded by a novel word order have found that older children tend to "correct" the novel word order to make it consistent with the language that they know (Abbot-Smith, Lieven, & Tomasello, 2001; Akhtar, 1999; Matthews, Lieven, Theakston, & Tomasello, 2005). One explanation is that learners implicitly assume that a different form should indicate a different meaning, since true synonymy is rare in language, both in morphology and in phrasal constructional patterns (e.g., Bolinger, 1977; Clark, 1987; Goldberg, 1995). While adults in the Wonnacott et al. (2008) study were willing to treat the artificial language learning context as providing pragmatic motivation for assigning a familiar meaning to one or more novel forms (cf. also Chang, Kobayashi, & Amano, 2009), in the current work we avoided potential complications posed by synonymous constructions by assigning a novel *function* to our novel form. In this case, it is clear that the target "correct" response is one that makes use of the novel word order.

Another benefit to studies that involve novel functions as well as novel forms is that, arguably, this is exactly the learning task that children face. They are not learning forms (or functions) in isolation but rather which formal patterns correspond to which abstract functions. There are a few studies that have taught children novel form-function pairings. Casenhiser and Goldberg (2005) exposed 6-year-old children to examples of a construction involving the novel form NP_1NP_2V and a novel abstract event semantics: the entity denoted by NP_1 (the *theme*) appeared in/on the location denoted by NP_2. The construction was presented in the context of English with novel verbs. For example, the sentence *The rabbit the hat moopos* referred to a scene in which a rabbit appeared on a hat. Children were exposed to the semantics by watching a set of 16 animated scenes accompanied by audio. The scenes were presented in a block with total exposure lasting approximately 3 min. The results demonstrated that there was better generalization of the abstract construction to novel vocabulary when the input was skewed such that half of the exemplars of the construction occurred with one particular nonsense verb, as opposed to equal numbers of exemplars with each novel verb (in each case the construction was presented the same number of times in total, and seen with the same total number of nonsense verbs – i.e. token and type frequency were held constant). Equivalent results were found with adult learners in related experiments (Goldberg, Casenhiser, & Sethuraman, 2004; cf. Boyd & Goldberg, 2009; Johnson & Goldberg, submitted for publication a; Year & Gordon, 2009, for limits on the advantage of skewed input). This result indicates that the extent to which learners generalize a construction is a function of the structure of the input and is thus generally in line with the findings of Wonnacott et al. (2008) and others (e.g., Gerken, 2006).

An additional factor beyond type of input may be the age of the learner. Boyd and Goldberg (2011) compared the learning of a novel construction in 5-year-olds, 7-year-olds, and adults. The novel construction again comprised an NP_1NP_2V form, although in this study the form was paired with approach events in which the entity denoted by NP_1 approached the entity denoted by NP_2 in the manner denoted by the verb. Again the exposure set comprised 16 sentences using the novel construction, and again there were five novel verbs. The distribution of verbs across sentences was skewed so that one verb occurred in half of the exposure sentences. One difference between this exposure set and that of the previous study was that each sentence involved the same two noun phrases (this point may be important—we return to it below).

After exposure, children were given a forced choice comprehension test. The novelty of test items was manipulated across three types of test items: *no novelty*, *novel verb*, and *all novel*. In no novelty test items, participants heard sentences that had been included in the exposure set, and distinguished the correct scene from a foil scene involving the same characters but an inappropriate semantics. In novel verb test items, participants heard an instance of the construction involving the two noun phrases from the exposure set but an unfamiliar verb, and again distinguished the correct scene (i.e., a scene depicting the characters in a novel approach event) from one with inappropriate semantics. In all novel test items, participants heard an instance of the construction comprised of entirely new vocabulary, and again distinguished the correct scene from one involving the same characters but inappropriate semantics. *Intransitive* test items were interspersed: these required participants to choose a non-approach scene to be sure that they were not simply preferring familiar scenes. The results demonstrated that adults and 7-year-olds showed no decrements in performance when confronted with increasingly novel test items. In contrast, 5-year-olds showed strong performance on the no novelty and intransitive test items, weaker performance on the novel verb items and weakest performance on the all novel items. This suggests that, unlike 7-year-olds and adults, 5-year-olds were relatively conservative in their learning: while they were able to extend the generalization to instances of the construction that were marginally similar to those they had been exposed to, their constructional representations were not general enough to allow extension to instances with no shared vocabulary. Note that, critically, Boyd and Goldberg were comparing generalization by these different age groups given precisely the same input. In addition to exploring the relationship between lexical conservatism and age, the study also further probed the types of representations children and adults acquired from their exposure to the construction. Note that for their no novelty, novel verb and all novel test items—and for all of the test items in Casenhiser and Goldberg (2005)—it would be possible for participants to distinguish correct and foil scenes given a *global* rather than precise knowledge of the semantics of the construction—i.e., recognizing that the NP_1NP_2V construction matched to approach semantics, but not necessarily linking NP_1 to the agent and NP_2 to the goal. To probe whether participants had acquired the more precise linking, Boyd and Goldberg also included additional *linking* test items. Here, participants heard familiar sentences (ones included in the exposure set) and saw two scenes, one that matched and another that was identical but with the agent and goal roles reversed. Adult and 7-year-old participants were able to distinguish the correct scenes from the foils; 5-year-olds were, however, at chance. Thus, in Boyd and Goldberg (2011) there were two ways in which 5-year-olds were more limited in their understanding of the novel construction than 7 year olds and adults: first, their understanding of the construction was more strongly tied to the lexical items with which it had been learned, and second, although they acquired the global semantics of the construction, they showed no learning of the precise mapping between word order and thematic roles. The current study sought to further probe both of these aspects of learning.

An important question is whether children of this age are constrained to begin with a period of lexically conservative learning regardless of the structure of the input, or whether there is a relationship between generalization and the structure of the input. One factor in the Boyd and Goldberg (2011) experiment that may have encouraged 5-year-olds to be particularly conservative was that every sentence in the input involved the same two noun phrases. This type of input actually provides little evidence for a construction which operates beyond this limited vocabulary. In line with this explanation, children performed significantly worse on test items where both the nouns and the verbs were novel (the all novel test items), than where the nouns were familiar and the verbs were novel (the novel verbs test items). (Note that there were no test items with familiar verbs and novel nouns).

The current study explores the process of novel construction learning in children with several goals in mind. First, we aimed to determine whether 5-year-olds's might show increasingly robust knowledge of the construction with more exposure. To this end, 3 days of input were included in the present exposure regime, whereas previous work with children had provided input on only a single day, and we compared learning at the end of the first and last day. Second, we sought to ascertain whether the structure of the input might play a role in constructional generalization to new lexical items. We chose to manipulate the type frequency of the construction since this type of manipulation has been established as important in promoting generalization across items in the domain of morphology (e.g. Bybee, 1985; Plunkett & Marchman, 1991). Specifically, we compared generalization to new verbs given two input sets: one in which the construction was demonstrated in the context of multiple verbs, and one in which the construction was demonstrated in the context of a single verb. Our hypothesis was that witnessing the construction with a single verb would result in less generalization of the construction to new verbs. Third, because the question of whether or not generalization has occurred may depend on the type of measurement used, we included three different tasks to assess acquisition of the novel construction: an act-out task, a production task, and a forced-choice comprehension task involving linking trials.

Experiment 1

Methods

Participants

Forty-two children (25 males, 17 females; age range 4:3–5:11; $M = 5:3$) were recruited from local state-funded primary schools in the Oxfordshire area, and from a database of parents who had previously expressed interest in taking part in studies. All children were raised as monolingual English speakers, and written consent to participate was received from all parents. An additional two children did not complete the procedure; their data are not included. Children were pseudo-randomly assigned to participate in either the *one verb* or *four verbs* condition (described below) such that age and gender were evenly distributed across conditions.

The novel construction

The construction that children were exposed to had both a novel form and a novel meaning—i.e., there is no equivalent construction in English.

- Form: verb-*ing* $noun_1$ $noun_2$ (VN_1N_2)
- Meaning: An $agent_2$ approaches a $goal_1$ in the manner denoted by the verb.

All exemplars of the construction consisted of a novel verb affixed with the high frequency, perceptually salient English suffix *-ing*, followed by two bare English nouns. Novel verbs were used so that children's interpretation of the construction would not be influenced by the semantics of known verbs, and the *-ing* suffix served to provide a strong cue that the novel words that occurred in the construction's first slot were, in fact, verbs.

A total of six English nouns were used—*bear, dog, frog, giraffe, gorilla*, and *rabbit*—each of which referred to one of six toy animals. Sentences were created by combining nouns with one of 14 novel verbs, each of which denoted a different manner of approach. For example, the sentence *Chadding rabbit gorilla* refers to an event in which a gorilla hops on its head towards a rabbit. Likewise, *Gaking frog bear* describes an event in which a bear slides towards a frog on its stomach. A complete set of the novel verbs used and their associated manners of approach is given in Appendix A.

Exposure

Children were given the opportunity to learn about the form and meaning of the novel construction by observing how the experimenter used 16 different exemplars of the construction. This group of sentences is referred to as the *exposure set*. In a typical exposure trial, the experimenter first produced a sentence from the exposure set by reading it aloud from a list. She then demonstrated its meaning using the toy animals. If, for example, the sentence was *Chadding rabbit gorilla*, the experimenter would make the gorilla hop towards the rabbit on its head. After the sentence's meaning had been enacted, the child was asked to repeat the sentence aloud. If the child repeated the sentence incorrectly, the experimenter would produce it again and give the child a single new opportunity to repeat it.[1]

Different 16-sentence exposure sets were generated for different experimental conditions, as described below. For each child's set, the verbs that were included were chosen randomly as a means of ensuring that specific verbs did not drive the recorded pattern of results. All six nouns were represented in each exposure set.

Design and predictions

Experiment 1 manipulated three variables—input group, verb familiarity, and day—in $2 \times 2 \times 2$ design. Children in the two input groups were exposed to different numbers of verb types in their exposure sets. In the one verb group, all 16 sentences in the exposure set featured the same verb. In the four verbs group, however, four different verbs appeared in the exposure set four times each. The intent of this manipulation was to engender the acquisition of constructional representations in the four verbs group that were more abstract—specifically at the verb position—than those of the one verb group.

Verb familiarity was manipulated within participants, and refers to the familiarity of verbs at *test*: half of all test trials made use of *familiar* nonsense verbs—i.e., the verbs that children had seen before as part of their exposure set—while the other half made use of *unfamiliar* nonsense verbs that were completely novel to children. Because the ability to deal correctly with unfamiliar items requires generalization, and the input group manipulation was specifically designed to encourage the acquisition of more general constructional representations in the four verbs group than the one verb group, it was hypothesized that children in the four verbs group would show relatively small decrements in performance on test trials featuring unfamiliar verbs relative to those featuring familiar verbs. Children in the one verb group on the other hand, were expected to show significantly larger decrements when encountering trials with unfamiliar verbs. We anticipated that this pattern would show up as a statistically significant interaction of input group and verb familiarity.

As described in more detail below, the heart of the experimental procedure covered 3 days, with all children tested on both *day one* and *day three*. Given that the amount of experience that children had with the novel construction was cumulative over days, improved performance was expected at test on day three relative to day one.

Because children in the one verb group had only one verb in their exposure whereas children in the four verbs group had four, a single verb from each child's exposure set was used in all familiar verb test trials. This ensured

[1] In addition, half of the children in each condition not only copied the sentence aloud but also used the toys to copy the action (e.g. make the gorilla hop on his head to the rabbit). Our initial analyses looked for main effects of this manipulation (we predicted that enacting the event would help learning of the linking between the event structure and form) and found none. To simplify presentation, we do not include mention of this factor in the remainder of the manuscript.

equivalent testing across the two input groups. An additional six verbs that were not seen during exposure were set aside for use in trials involving unfamiliar verbs—one for each of three different tasks (described below) on the two test days (day one and day three). We anticipated and controlled for the possibility that idiosyncratic features of certain verbs might impact our results by having each child receive a random assignment of exposure and test verbs. This ensured that which verbs were familiar at test and which were unfamiliar varied across children.

In addition to randomizing nonsense verb assignment, we controlled test items across the within-participant conditions by using the same unique combinations of nouns across familiar and unfamiliar test items,[2] and using identical test items on day one and day three (apart from the use of different unfamiliar verbs).

Tasks

Three different tasks were used to measure what children had learned from exposure to the novel construction: *act out comprehension*, *production*, and *forced choice comprehension of linking*. In the act out task, children were told that they would have a chance to show what various sentences meant by moving the animals themselves. For act out trials involving familiar verbs, the experimenter would read a test sentence to a child, then ask the child to enact its meaning using the same set of toy animals used during exposure. For act out trials involving unfamiliar verbs the procedure was identical, with the exception that the meaning of the unfamiliar verb was first illustrated to children. This was accomplished in three steps. First, the experimenter would remind children of a sentence seen during exposure (e.g., by saying ''Remember 'Chadding rabbit gorilla'?'' while enacting its meaning). Then, the experimenter would use a different animal to show that the verb from the exposure sentence could be used intransitively (e.g., the experimenter would say ''Now this is just chadding,'' while making the animal hop back and forth on its head). Finally, the unfamiliar verb would be introduced in an intransitive structure with yet another animal (e.g., the experimenter would say, ''This is ralling,'' while showing a bee spinning forward). After illustrating the unfamiliar verb's meaning in this manner, the experimenter proceeded by asking the child to enact different test sentences containing the verb, as in trials with familiar verbs. All act out trials were videotaped for later coding and analysis.

In the production task, children were told that they would now be producing their own sentences themselves, but that the experimenter would help them by giving them the first word. For each trial, the experimenter enacted an approach event and began a description of it by producing the verb. Children then completed the experimenter's description however they saw fit. The procedure for trials involving familiar and unfamiliar verbs was the same, with the exception that in trials with unfamiliar verbs children were warned that they would be seeing actions that they had no prior experience with. All production trials were videotaped for later coding and analysis.

The forced choice task was specifically designed to test children's knowledge of the novel construction's linking rules. Each forced choice trial contained two movies, which appeared side by side on a computer. The movies played in turn, and then played together along with a pre-recorded exemplar of the novel construction that matched the content of only one movie. Children were asked to choose the movie that matched the sentence they heard by touching it on the computer screen. Responses were recorded by the experimenter using a mouse. Crucially, the movies in each trial demonstrated approach events in which the agent and goal roles were reversed. For example, if the audio was *Chadding rabbit gorilla,* then the correct match was to the movie in which a gorilla approached a rabbit, and the incorrect match was to the movie in which a rabbit approached a gorilla. The procedure for trials involving familiar and unfamiliar verbs was identical, except that for unfamiliar verbs children were warned that they would be hearing words that they had not heard before.

Procedure

All children were tested individually, and each child received a certificate thanking them for their help.

The experiment unfolded over the course of 4 days (days zero, one, two, and three), beginning on day zero with an orientation session in which children were introduced to the toy animals that would be used in the experiment and were asked to name them. Children were generally very successful at this. Where they produced a different label from the one used in the exposure sentences (e.g., saying ''teddy'' instead of ''bear''), they were given the appropriate label and asked to repeat it (e.g., ''He likes to be called 'bear.' Can you say 'bear'?'').

Day one of the experiment consisted of a vocabulary check, then an exposure block, then a test block. The vocabulary check was performed to ascertain whether children remembered the labels given to each animal on day zero. As on day zero, feedback was provided when children offered an incorrect label. The exposure block began with an introduction in which children were told that they were going to learn how to talk like the animals, who said things a bit differently from us. They were then given 16 exposure trials, one for each of the sentences in their exposure set. The test block consisted of 12 trials using a single familiar verb (four act out trials, then four production trials, then four forced choice trials), followed by 12 trials using three different unfamiliar verbs (four act out trials with one unfamiliar verb, then four production trials with a new unfamiliar verb, then four forced choice trials with yet another unfamiliar verb).

Day two of the experiment consisted of exposure only; no testing was done. Children worked through their exposure sets twice, for a total of 32 exposure trials. Short breaks were taken every eight trials.

Exposure and testing on day three were exactly the same as on day one, with the exception that different new verbs were used in the unfamiliar verb tests.

[2] In addition, for familiar verb test items, half of the sentences had actually occurred in the exposure set and half had not. Our initial analyses looked for main effects of this manipulation (predicting that children might show stronger performance with sentences that had occurred in their input), but found none. To simplify presentation, we do not include mention of this factor in the remainder of the paper.

Please cite this article in press as: Wonnacott, E., et al. Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language* (2012), doi:10.1016/j.jml.2011.11.004

The order in which exposure and test sentences were presented was randomized across participants, subject to the constraints already described.

Results

All data were analyzed in the R computing environment (R Development Core Team, 2010) using logit mixed models (Baayen, 2008; Jaeger, 2008; Quené & van den Bergh, 2008). We used a backward stepwise model comparison procedure starting with models with maximally-specified fixed and random effects structures to test whether individual fixed and random parameters were justified by the data.[3] In addition, fixed parameters reflecting a priori hypotheses (i.e., the input group × verb familiarity interaction and its constitutive terms, along with the day effect) were included in all models in which they varied. Reported models containing large numbers of parameters are summarized in tables for ease of presentation; simpler models are described in full in the text. Note that while we refer to proportion results in many places throughout the text, this is done predominantly to facilitate reader interpretation; unless otherwise noted, all statistical analyses were conducted in log-odds space.

Below, we address the outcome of each of the three experimental tasks in turn: act out comprehension, production, and forced choice comprehension of linking. There were no correlations between children's precise ages and any of the measures discussed below.

Act out comprehension results

Recall that in the act out task children heard sentences of the form VNN (e.g., *Chadding rabbit gorilla*), and were asked to use toy animals to enact the events described by the sentences. For this task, generating a response that is entirely veridical to the input requires learning at least two pieces of information. First, children must acquire *construction semantics*: do they know that VNN forms encode approach events? Second, children must learn *verb semantics*: have they acquired the mappings between specific verbs and specific manners of approach? Our central interest in this work is on the acquisition of construction semantics. In the present section we thus consider whether or not children produced approach actions, independently of whether they applied the correct manner of approach. Verb semantics, and the relationship between the acquisition of verb semantics and the acquisition of construction semantics, is considered in a separate section below.

All trials were coded into one of three categories: *approach correct*, *approach incorrect*, and *other*. In trials coded as approach correct, children produced an approach action with the correct linking rules (i.e., N_2 approached N_1 in some manner). In trials coded as approach incorrect, children produced an approach action with incorrect linking rules (i.e., N_1 approached N_2 in some manner). All remaining trials were coded as other. The other category included responses in which children moved both animals, moved only one animal, acted out a transitive action with the animals, or did not move either animal. A numerical breakdown of the different types of "other" responses is given in Appendix B. All trials were coded initially by the third author; a 10% sample was then independently coded by a research assistant. Inter-coder reliability for this sample was found to be 98%. Table 1 gives the proportion of response types in each of the experimental conditions; Fig. 1 summarizes these data graphically. A recording error led to the loss of act out data for one child. As a result, the reported means and subsequent statistical analyses cover 41 children rather than 42.

Children's acquisition of construction semantics was assessed using two types of analyses. The first—which we refer to as the *global* analysis—sought to determine whether the manipulated variables affected children's acquisition of the overall semantics of the novel construction. That is, did children learn that the construction referred to approach events generally, regardless of whether they acquired the correct mapping of nominal arguments to thematic roles? The second analysis—the *linking* analysis—dealt with the specifics of the syntax–semantics mapping. Assuming that the construction's global semantics had been learned, did children know that N_1 mapped to the goal argument, and N_2 to the agent?

Act out global analysis. Note that both approach correct and approach incorrect responses demonstrate that children have learned that the construction generally denotes approach events. These are represented by dark and light gray portions of the bars in Fig. 1, which show that approach events (represented by the sum of dark and light gray) were enacted at an overall rate of 74%. To determine the effect that the experimental manipulations had on children's ability to learn the general association between VNN forms and approach semantics, a mixed model was fit to the act out data using the likelihood of an approach response as the dependent variable. This model is summarized in Table 2.

The model results show that children in the four verbs group were marginally more likely than children in the one verb group to enact approach responses ($\beta = 6.85$, $z = 1.74$, $p = .082$; four verbs: 82%; one verb: 65%), and that children were significantly less likely to enact approach responses on trials involving unfamiliar versus familiar verbs ($\beta = -8.12$, $z = -2.18$, $p = .029$; unfamiliar: 60%; familiar: 87%). Moreover, both of the predictions outlined in the methods section regarding the effects of input on construction learning were borne out. Specifically, as children were exposed to additional exemplars of the novel construction across days, they were increasingly likely to associate VNN forms with approach semantics, as evidenced by the significant effect of day ($\beta = 5.46$, $z = 2.24$, $p = .025$; day one: 68%; day three: 79%). Additionally, the significant interaction of input group and verb familiarity ($\beta = 14.82$, $z = 1.99$, $p = .046$) indicates that children's ability to extend the construction to unfamiliar verbs was not equivalent across the two groups. As Fig. 1 illustrates, children in the one verb group showed larger decrements in performance on trials involving unfamiliar verbs than children in the

[3] Alternative analyses relying on ANOVAs and logit mixed models in which all random effects were included—regardless of whether model comparison indicated that they were justified by the data—yielded results that were nearly indistinguishable from those reported.

Table 1
Proportion of response types (*SE*) by condition in the Experiment 1 act out comprehension task.

	Day one			Day three		
	Approach correct	Approach incorrect	Other	Approach Correct	Approach incorrect	Other
One verb						
Familiar	0.50 (0.09)	0.30 (0.07)	0.20 (0.09)	0.58 (0.08)	0.38 (0.08)	0.05 (0.05)
Unfamiliar	0.20 (0.08)	0.21 (0.07)	0.59 (0.10)	0.24 (0.08)	0.21 (0.07)	0.55 (0.11)
Four verbs						
Familiar	0.33 (0.06)	0.44 (0.07)	0.23 (0.08)	0.58 (0.09)	0.36 (0.09)	0.06 (0.05)
Unfamiliar	0.43 (0.09)	0.30 (0.09)	0.27 (0.09)	0.59 (0.09)	0.25 (0.08)	0.16 (0.08)

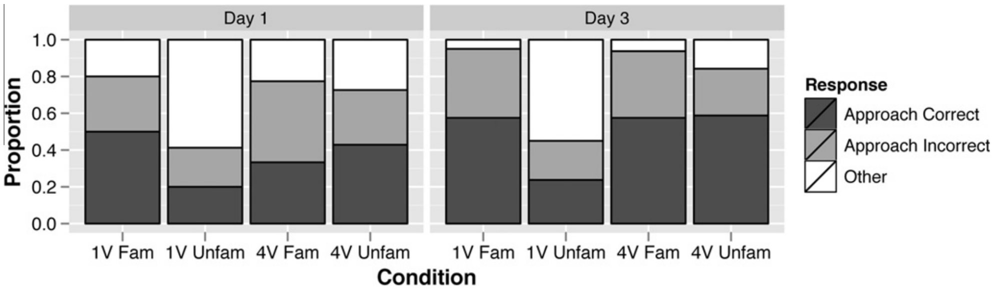

Fig. 1. Proportion of response types by condition in the Experiment 1 act out comprehension task. For error bars please refer to the standard errors in Table 1. *1V* = one verb; *4V* = four verbs; *Fam* = familiar; *Unfam* = unfamiliar.

Table 2
Mixed model of the likelihood of an approach response in the Experiment 1 act out comprehension task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	7.69	2.00	3.85	.00012
Input Group = four verbs	6.85	3.93	1.74	.082
Verb Familiarity = unfamiliar	−8.12	3.73	−2.18	.029
Day = day three	5.46	2.44	2.24	.025
Input Group × Verb Familiarity	14.82	7.44	1.99	.046
Random effects				
	SD			
Participant (intercept)	7.69			
Verb Familiarity	13.93			
Day	10.56			

four verbs group. This outcome is consistent with the acquisition of more general constructional representations in the four verbs group based on exposure to input that exhibited variability at the verb slot.

Act out linking analysis. While the global analysis indicates that input-dependent learning of the general association between VNN forms and approach semantics occurred in Experiment 1, it does not address whether children acquired any knowledge of the novel construction's linking rules. There are a number of ways this question could be answered. In subsequent sections (see Experiment 2) we investigate whether exposure to the novel construction results in a shift in linking preferences relative to a control group. The strategy here, however, is to use the presence of day effects to establish that linking rules are being learned. The logic of this kind of analysis is as follows: if the likelihood of producing an approach correct response goes up from day one to day three, while the likelihood of producing an approach incorrect response either stays the same or decreases, then we can conclude that the additional exposure to the construction that was provided on days two and three helped children to hone in on the correct linking pattern.

Note that we are not seeking to prove here that children's knowledge of the novel construction's linking rules is particularly good. As is evident from the relatively sizeable portion of approach incorrect responses in Fig. 1, quite a bit of improvement would be needed before one could say that children were manifesting the kind of deterministic behavior that is characteristic of adult-like knowledge of linking rules. Instead, our goal is simply to determine whether there has been significant, if incremental, input-dependent learning of the mapping from thematic roles to syntactic positions.

Two mixed models were fit to the act out data to test for such learning. The first used the likelihood of an approach correct response as the dependent variable (the dark gray portions of the bars in Fig. 1); the second used the likelihood of an approach incorrect response as the dependent (the light gray portions of the bars in Fig. 1). These models are summarized in Tables 3 and 4.

The fixed effects structure of the model in Table 3 closely resembles that of the model reported in Table 2 for the global act out analysis. The likelihood of an approach correct response was higher in the four verbs group than the one verb group ($\beta = 1.33$, $z = 1.98$, $p = .047$; four verbs: 47%; one verb: 38%), and was lower on trials involving unfamiliar verbs ($\beta = -1.74$, $z = -4.20$, $p < .0001$; unfamiliar: 36%; familiar: 49%). Further, as in the global act out analysis, the familiarity effect differed across the two

Table 3
Mixed model of the likelihood of an approach correct response in the Experiment 1 act out comprehension task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	−0.76	0.34	−2.23	.026
Input Group = four verbs	1.33	0.67	1.98	.047
Verb Familiarity = unfamiliar	−1.74	0.41	−4.20	<.0001
Day = day three	1.07	0.40	2.69	.0072
Input Group × Verb Familiarity	3.93	0.81	4.84	<.0001
Random effects				
	SD			
Participant (intercept)	1.98			
Verb Familiarity	1.75			
Day	2.37			
Verb Familiarity × Day	3.81			

Table 4
Mixed model of the likelihood of an approach incorrect response in the Experiment 1 act out comprehension task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	−1.70	0.27	−6.18	<.0001
Input Group = four verbs	0.70	0.52	1.36	.18
Verb Familiarity = unfamiliar	−1.82	0.42	−4.29	<.0001
Day = day three	−0.16	0.40	−0.40	.69
Input Group × Verb Familiarity	0.46	0.83	0.56	.58
Random effects				
	SD			
Participant (intercept)	1.55			
Verb Familiarity	2.00			
Day	2.43			
Verb Familiarity × Day	3.89			

groups ($\beta = 3.93$, $z = 4.84$, $p < .0001$): children in the one verb group showed larger familiarity effects than children in the four verbs group. Of primary importance though, is the statistically significant effect of day. The likelihood of producing an approach correct response increased from days one to three ($\beta = 1.07$, $z = 2.69$, $p = .0072$; day one: 37%; day three: 49%).

In contrast, when the data were used to predict the likelihood of an approach incorrect response (see Table 4), the fitted model showed a significant effects of verb familiarity ($\beta = -1.82$, $z = -4.29$, $p < .0001$; unfamiliar: 25%; familiar: 37%), but no input group effect ($\beta = 0.70$, $z = 1.36$, $p = .18$, and crucially, no day effect ($\beta = -0.16$, $z = -0.40$, $p = .69$. The null day result, combined with the statistically significant increase in approach correct responses across days in Table 3 model, is consistent with the acquisition of linking rules from the input. Moreover, it indicates that the day effect reported in the global act out analysis (see Table 2) was not driven simply by increases in the production of approach actions, without regard for whether or not they exemplified knowledge of the construction's linking rules, but specifically by increases in the production of approach actions with the *correct* linking. Thus, as children gained additional experience with the novel construction, they were increasingly likely to interpret N_2 as the agent in an approach event, and N_1 as the goal.

Verb semantics. The previous analyses considered children's understanding of construction semantics—that is, whether they knew anything about the global relationship between VNN forms and approach semantics, and whether there was any evidence that they had learned about the novel construction's linking rules. Here we focus on verb semantics by asking two questions. First, did the type of input that children were exposed to affect their ability to correctly replicate the different manners of approach encoded by verbs in the act out task? And second, is there any relationship on individual test trials between children's ability to replicate verb manners, and their ability to map VNN forms to approach semantics? Coding for verb manner was done independently by the third author and a research assistant. Trials for which the verb manner that a child enacted matched the meaning of the verb given in Appendix A were coded as correct; all other trials were coded as incorrect. Trials on which the two coders disagreed, or for which the manner could not be determined from the video were excluded from analysis. This amounted to less than 4% of the data. Table 5 shows the mean proportion of correct verb manner enactments across conditions.

The manner data were analyzed in the mixed model summarized in Table 6, which used the likelihood of a correct response as the dependent variable. The model indicates that children in the four verbs group were significantly less likely to enact the correct verb manner than children in the one verb group ($\beta = -6.18$, $z = -4.07$, $p < .0001$; four verbs: 53%; one verb: 89%), that the likelihood of a correct enactment was higher on day three than on day one ($\beta = 2.84$, $z = 2.91$, $p < .0037$; day three: 78%; day one: 64%), and that verb familiarity acted differently across the one verb and four verbs groups ($\beta = 9.01$, $z = 3.01$, $p < .0026$).

As a means of following up on the statistically significant interaction, separate non-parametric conditional inference tests (Strobl, Malley, & Tutz, 2009) were conducted over the familiar and unfamiliar verb data, collapsing across days. These found that while children in the four verbs group were significantly worse than children in the one verb group at enacting the manners of familiar verbs ($p < .001$), there was no difference between groups when they were tested on unfamiliar verbs (n.s.). This pattern is straightforwardly related to the experimental design and procedure. Enacting the correct verb manner on familiar verb trials is much more difficult in the four verbs group, since children there must encode and recall four times as many verb-manner associations as children in the one verb group. The lack of a group difference on

Table 5
Proportion of correct verb manner enactments (*SE*) by condition in the Experiment 1 act out task.

	Day one	Day three
One verb		
Familiar	0.94 (0.05)	0.95 (0.05)
Unfamiliar	0.79 (0.08)	0.87 (0.07)
Four verbs		
Familiar	0.18 (0.06)	0.46 (0.09)
Unfamiliar	0.68 (0.10)	0.80 (0.08)

Table 6
Mixed model of the likelihood of a correct verb manner enactment in the Experiment 1 act out comprehension task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	3.96	0.81	4.92	<.0001
Input Group = four verbs	−6.18	1.52	−4.07	<.0001
Verb Familiarity = unfamiliar	2.31	1.50	1.54	.12
Day = day three	2.84	0.98	2.91	.0037
Input Group × Verb Familiarity	9.01	2.99	3.01	.0026
Random effects				
	SD			
Participant (intercept)	3.24			
Verb Familiarity	5.97			
Day	4.27			

unfamiliar verb trials is due to the fact that the experimenter demonstrated the meaning of all unfamiliar verbs at the beginning of the test. Children thus only needed to retain the new verb-manner association across the four test trials, rather than recalling it from the training set, regardless of whether they participated in the one verb group or four verbs group.

Our second question relating to verb semantics was whether getting the verb manner correct on individual trials might correlate with children's ability to map from VNN forms to approach semantics. Addressing this issue involved attempting a replication of the analysis of the likelihood of an approach response summarized in Table 2, but with two changes. First, because the main result from Table 2 model—i.e., the statistically significant interaction of input group and verb familiarity—was constant across days, we collapsed across days in the current analysis. Second, verb manner was added as a predictor and its interaction with input group and verb familiarity was tested.

The data are summarized in Table 7 and Fig. 2. Note that the addition of verb manner as a predictor led to a highly unbalanced design. For example, because only two children from the one verb group incorrectly enacted verb manners on familiar verb trials, there are only two participants contributing to that cell of the design. One of the children in this condition always enacted approach events and the other never did. This accounts for the very large error bar depicted for the condition in Fig. 2.

A consequence of the unbalanced dataset was that we were unable to fit a logit mixed model. Most models including a three-way interaction parameter failed to converge, which made it impossible to identify the model that provided the optimal balance between fit and complexity. This motivated the use of non-parametric conditional inference tests to look for pairwise between-groups differences (Strobl et al., 2009). The conditional inference analysis revealed that the probability of an approach response did not differ across the one verb and four verbs groups when the test verb was familiar and the verb manner was correctly enacted, or when the test verb was familiar and the verb manner was incorrectly enacted, or when the test verb was unfamiliar and the verb manner was incorrectly enacted (all *p*-values >.05). When the test verb was unfamiliar and the verb manner was correctly enacted, however, children in the four verbs group were significantly more likely to produce an approach response than children in the one verbs group, $p = .001$ (compare the two lighter bars on the right side of Fig. 2).

This pattern across the four conditional inference tests is consistent with a three-way interaction of input group, verb familiarity, and verb manner. Further, it suggests that the statistically significant two-way interaction of input group and verb familiarity reported in Table 2 (i.e. the type frequency effect) is driven primarily by performance on act out trials in which verb manners were correctly enacted. In other words, low productions of approach actions in trials where children do not enact the correct action reflect more general uncertainty (regardless of condition), whereas low productions of approach actions in trials where the verb manners were correctly enacted reflect a response to the input structure (hence the difference across conditions).

Summary of act out results. Across conditions children showed an ability to associate the novel VNN form with approach semantics, with even the weakest condition enacting approach events on 41% of trials. The type frequency of verbs in the input did, however, affect children's tendency to extend approach semantics to unfamiliar verbs. When children were exposed to sentences with four different verbs, the majority of their responses conveyed approach semantics, regardless of the familiarity of the verbs used at test. In contrast, when their exposure contained only one verb, children were markedly less likely to produce events with approach semantics when the test verb was unfamiliar than when it was familiar. In this condition, instead of producing approach actions, children tended to produce symmetrical actions in which both animals were moved equally—i.e. they settled on a two-participant interpretation of the novel construction that did not incorporate approach semantics. This suggests that the number of verbs included in the exposure set significantly affected whether children abstracted a verb-independent constructional representation—i.e., the type frequency of the construction affects generalization.

Unsurprisingly, the input also affected the learning of familiar verb semantics: children in the one verb group had seen the verb-manner association four times more frequently in their input than each of the associations which the four verbs group had to learn, and they showed better learning of that association. For unfamiliar verbs, children in the four verb and one verb group showed no difference in their ability to enact the verb meanings (reflecting the

Table 7
Proportion of approach enactments (*SE*) in the Experiment 1 act out task, with input group, verb familiarity, and verb manner as predictors.

	Familiar verbs	Unfamiliar verbs
Correct verb manner		
One verb	0.92 (0.04)	0.36 (0.09)
Four verbs	1.00 (0.00)	0.81 (0.08)
Incorrect verb manner		
One verb	0.50 (0.50)	0.55 (0.20)
Four verbs	0.83 (0.06)	0.63 (0.13)

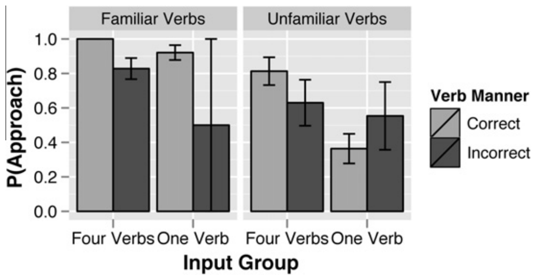

Fig. 2. Proportion of approach responses in the Experiment 1 act out task, with input group, verb familiarity, and verb manner as predictors. Error bars show the standard error of the mean.

fact that unfamiliar verbs are introduced during the test), however here conditional inference tests revealed a relationship between correct enactment of verb meanings and the association of VNN forms with approach semantics. The four verbs group showed a significant advantage over the one verb group on trials involving unfamiliar verbs, but only when verb manners had been correctly enacted, i.e. the type frequency effect for construction learning is apparent only for trials where children enact the correct verb meaning. This suggests that in trials where children do *not* enact the correct verb meaning, poor performance on construction semantics reflects more general uncertainty, rather than a response to input structure. In trials where they do enact the correct semantics, poor performance on construction semantics reflects the influence of the input.

Finally, although children's use of linking rules was not adult-like, the fact that they showed significant improvement across days indicates incremental learning of the mapping from syntactic categories to argument structure.

Production results

Recall that in the production task children witnessed an event in which one animal approached another in some manner (e.g., a bear jumped towards a frog), and heard the first word of a sentence describing the event (e.g., ''Teeping…''). Their job was to complete the sentence.

Each response was assigned one of three codes that mirrored the codes used for the act out task: *VNN correct*, *VNN incorrect*, and *other*. Responses that were categorized as VNN correct consisted of the label for the event's goal animal followed by the label for its agent animal—in other words, the utterance correctly exemplified the novel construction's linking rules. VNN incorrect responses used the same labels, but in the opposite (incorrect) order. The other code was assigned to all remaining responses. This included utterances in which only the goal or the agent argument was produced, responses involving the wrong labels, responses that consisted of English or pseudo-English sentences, and trials in which children failed to give any response. A numerical breakdown of responses in the other category is provided in Appendix C. Note that whether or not children repeated the initial verb did not affect coding. For instance, both ''Teeping frog bear'' and ''frog bear'' would be coded as VNN correct given a scene in which a bear jumped towards a frog and an experimenter prompt of ''Teeping…''. Table 8 gives the mean proportion of response types in each of the experimental conditions. These data are summarized graphically in Fig. 3.

As a means of assessing children's acquisition of the novel construction, the production data were subjected to the same kinds of analyses used above to evaluate the act out data. As before, the global analysis appraised learning of the general relationship between approach semantics and VNN forms, without regard for nominal ordering, while the linking analysis sought to determine whether children had learned about the mapping of agents in approach events to N_2, and goals to N_1.

Production global analysis. Both VNN correct and VNN incorrect responses are consistent with children having learned that approach events are encoded as a verb followed by two nouns. These response types are represented by the dark and light gray portions of the bars in Fig. 3, which show that approach events were described using VNN forms (represented by the sum of dark and light gray) at an overall rate of 72%. To determine the effect that the experimental manipulations had on the production of VNN forms, a mixed model was fit to the production data using the likelihood of a VNN response as the dependent variable. This model is summarized in Table 9.

Table 9 model has much in common with its act out counterpart in Table 2. In fact, the only difference between the two global analyses in terms of fixed effects is that the production data fail to demonstrate a significant effect of verb familiarity ($\beta = 0.25$, $z = 0.17$, $p = .86$). All other fixed effects reach similar levels of significance. For instance, the production dataset shows that children were marginally more likely produce a VNN response in the four verbs group than in the one verb group ($\beta = 4.04$, $z = 1.83$, $p = .67$; four verbs: 81%; one verb: 63%). And most importantly, the production and act out data agree in terms of the effects that indicate input-based learning of the novel construction. The production data show improvement in the general mapping from approach semantics to VNN forms across days ($\beta = 1.31$, $z = 2.47$, $p = .014$; day one: 69%; day three: 76%). Further, there was no difference across tasks in the unevenness of the familiarity effect: as can be seen in Fig. 3, verb familiarity had a larger effect on the likelihood of a VNN production in the one verb group than the four verbs group ($\beta = 6.20$, $z = 2.14$, $p = .033$). These results indicate, as expected, that increased exposure across days

Table 8
Proportion of response types (*SE*) by condition in the Experiment 1 production task.

	Day one			Day three		
	VNN correct	VNN incorrect	Other	VNN correct	VNN incorrect	Other
One verb						
Familiar	0.35 (0.09)	0.41 (0.08)	0.24 (0.08)	0.43 (0.10)	0.38 (0.09)	0.20 (0.07)
Unfamiliar	0.18 (0.08)	0.26 (0.09)	0.56 (0.11)	0.29 (0.08)	0.23 (0.07)	0.49 (0.11)
Four verbs						
Familiar	0.42 (0.09)	0.35 (0.08)	0.23 (0.07)	0.57 (0.08)	0.30 (0.07)	0.13 (0.07)
Unfamiliar	0.38 (0.09)	0.39 (0.09)	0.24 (0.08)	0.49 (0.09)	0.36 (0.09)	0.15 (0.08)

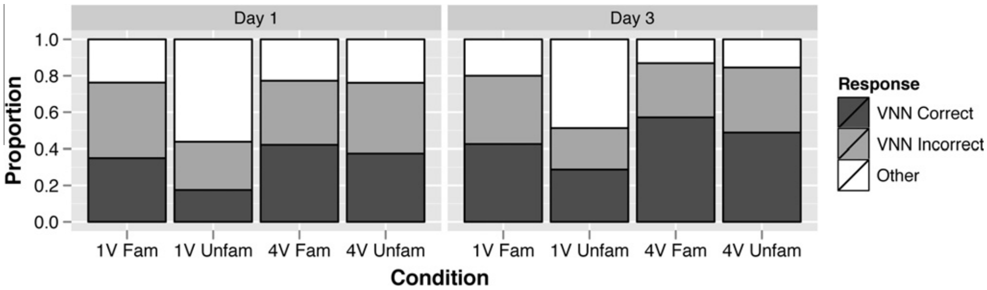

Fig. 3. Proportion of response types by condition in the Experiment 1 production task. For error bars please refer to the standard errors in Table 8. *1V* = one verb; *4V* = four verbs; *Fam* = familiar; *Unfam* = unfamiliar.

Table 9
Mixed model of the likelihood of a VNN response in the Experiment 1 production task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	3.27	1.11	2.95	.0032
Input Group = four verbs	4.04	2.20	1.83	.067
Verb Familiarity = unfamiliar	0.25	1.45	0.17	.86
Day = day three	1.31	0.53	2.47	.014
Input Group × Verb Familiarity	6.20	2.90	2.14	.033
Random effects				
	SD			
Participant (intercept)	6.09			
Verb Familiarity	7.17			
Day	3.90			
Verb Familiarity × Day	6.01			

facilitated acquisition of the novel construction's general form-function association, and that exposure to more verb types in the four verbs group led to the acquisition of representations that were robust to changes in verb familiarity at test.

Production linking analysis. We now turn to the question of whether children showed learning of the precise linking between word order and thematic roles—i.e., whether they learned that goal arguments should be produced at N_1 and agents at N_2. As noted above in the act out analysis, there are a number of ways to test for linking rule acquisition. Here we focus on the presence of day effects. Specifically, if the likelihood of VNN correct responses goes up across days while the likelihood of VNN incorrect responses stays the same or decreases, this would be consistent with the hypothesis that additional input leads to improved knowledge of the novel construction's linking rules. Note again that we are not attempting to show that children's linking rule knowledge is mature—the large proportions of VNN incorrect responses in Fig. 3 definitively belie that notion. Instead, the goal is to establish whether there is evidence of incremental learning of linking rules with input.

Two models were fit to the production data, the first using the likelihood of a VNN correct response as the dependent (represented by the dark gray portions of the bars in Fig. 3), and the second using the likelihood of a VNN incorrect response as the dependent (represented by the light gray portions of the bars in Fig. 3). These models are summarized in Tables 10 and 11.

The analysis of the likelihood of VNN correct responses in Table 10 indicates that children produced more VNN correct responses in the four verbs group ($\beta = 1.58$, $z = 1.98$, $p = .048$; four verbs: 46%; one verb: 31%), fewer on trials involving unfamiliar test verbs ($\beta = -1.06$, $z = -3.33$, $p = .00088$; unfamiliar: 33%; familiar: 44%), and that there was indeed a significant increase in the likelihood of a VNN correct response from day one to day three ($\beta = 1.16$, $z = 2.94$, $p = .0033$; day one: 33%; day three: 45%). The analysis of the likelihood of VNN incorrect responses in Table 11 shows a similar effect of verb familiarity ($\beta = -0.93$, $z = -2.57$, $p = .010$; unfamiliar: 31%; familiar: 36%), and an interaction of familiarity and input group ($\beta = 1.76$, $z = 2.50$, $p = .012$), which indicates that the effect of familiarity on VNN incorrect responding was not equivalent across the one verb and four verbs groups. Crucially, the analysis also failed to find a statistically significant effect of day ($\beta = -0.22$, $z = -0.62$, $p = .53$). These results—a significant day effect in the VNN correct analysis and a null effect in the VNN incorrect analysis—suggest that the effect of day reported in the global production analysis (see Table 9) was due to an increase in VNN responses with the *correct* linking, rather than to a general increase in VNN responses.

Please cite this article in press as: Wonnacott, E., et al. Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language* (2012), doi:10.1016/j.jml.2011.11.004

Table 10
Mixed model of the likelihood of a VNN correct response in the Experiment 1 production task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	−1.01	0.41	−2.45	.014
Input Group = four verbs	1.58	0.80	1.98	.048
Verb Familiarity = unfamiliar	−1.06	0.32	−3.33	.00088
Day = day three	1.16	0.39	2.94	.0033
Input Group × Verb Familiarity	0.99	0.63	1.57	.12
Random effects				
	SD			
Participant (intercept)	2.49			
Verb Familiarity	1.24			
Day	1.86			

Table 11
Mixed model of the likelihood of a VNN incorrect response in the Experiment 1 production task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	−1.45	0.38	−3.80	.00014
Input Group = four verbs	0.21	0.75	0.28	.78
Verb Familiarity = unfamiliar	−0.93	0.36	−2.57	.010
Day = day three	−0.22	0.35	−0.62	.53
Input Group × Verb Familiarity	1.76	0.70	2.50	.012
Random effects				
	SD			
Participant (intercept)	2.39			
Verb Familiarity	1.69			
Day	2.37			
Verb Familiarity × Day	2.93			

As in the act out data, this pattern of results is consistent with improved knowledge of the novel construction's linking rules as a result of increased exposure.

Summary of production results. Children's behavior in the production task closely mirrored their behavior in the act out task. The global analysis results from both datasets demonstrate that children are more ready to generalize the novel construction across verbs in the four verbs group than the one verb group. And the outcome of both linking analyses indicates that linking rule performance improved across days with additional input.

Forced choice comprehension of linking results

Recall that in each forced choice trial children heard an exemplar of the novel construction and were asked to choose between a movie that accurately depicted its meaning, and one that also showed an approach event, but with the participants' roles reversed. This test therefore probes whether children have learned the specific linking rules associated with the novel construction. The forced choice data are summarized in Table 12 and Fig. 4, which show the mean proportion of correct responses across conditions.

Our analysis of the forced choice data is summarized in the mixed model results reported in Table 13. No model estimates were significant. This includes the intercept ($\beta = 0.11$, $z = 0.69$, $p = .49$), which indicates that the measured likelihood of picking the correct movie (i.e., 52%) was no different than chance. Should we conclude from this null result that children learned nothing about the novel construction's linking rules? Not necessarily. First, recall that the linking analyses of the act out and production data have already provided positive evidence that linking performance improves with additional input, although that result was not replicated in this third task. Second, although the forced choice task had only two alternatives, it does not follow that children were equally likely to choose one or the other as the meaning of the test sentences they heard. If, for example, they came into the experiment with an a priori bias to interpret the first noun of NN sequences as an agent, then 52% correct might represent a substantial amount of learning. We return to this point in more detail in the Experiment 1 discussion section.

Behavior across the three tasks

If the act out, production, and forced choice tasks tap into the same knowledge, then children's performance should be correlated across tasks. To test for this possibility we calculated five measures for each child, collapsing across verb familiarity and day. Two of these—the probability of an approach response in the act out data, and the probability of a VNN response in the production data—assessed acquisition of the general relationship between approach semantics and VNN forms. The other three measures considered the extent to which the construction's linking rules had been learned: the probability of an approach correct response in the act out data (calculated as the number of approach correct responses out of all approach responses), the probability of a VNN correct response in the production data (calculated as the number of VNN correct responses out of all VNN responses), and the probability of a correct response in the forced choice data.

Fig. 5 illustrates the relationships that exist between these measures, broken down by input group. Each panel represents a different pairwise correlation. Note that all regression lines have positive slopes, but that across all panels the slopes are steeper for the four verbs group than the one verb group, suggesting stronger relationships across tasks in that condition. This observation is borne out in statistical analysis of the data. The four verbs correlations are significant in all panels (A: $r^2 = .28$, $p = .014$; B: $r^2 = .38$, $p = .048$; C: $r^2 = .70$, $p < .0001$; D: $r^2 = .52$, $p = .00032$). In contrast, while the one verb correlations all show relationships in the same direction as the four verbs group, some are statistically significant (B: $r^2 = .35$, $p = .0074$), others are marginally significant (C: $r^2 = .15$, $p = .10$), and others are not

Table 12
Proportion of response types (*SE*) by condition in the Experiment 1 forced choice comprehension task.

	Day one	Day three
One verb		
Familiar	0.53 (0.07)	0.51 (0.08)
Unfamiliar	0.50 (0.07)	0.48 (0.07)
Four verbs		
Familiar	0.53 (0.08)	0.54 (0.07)
Unfamiliar	0.49 (0.08)	0.58 (0.08)

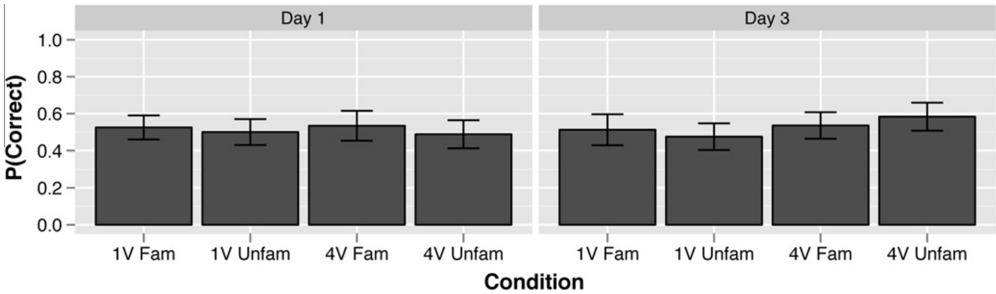

Fig. 4. Proportion of response types by condition in the Experiment 1 forced choice comprehension task. Error bars show the standard error of the mean. *1V* = one verb; *4V* = four verbs; *Fam* = familiar; *Unfam* = unfamiliar.

Table 13
Mixed model of the likelihood of a correct response in the Experiment 1 forced choice task.

	Estimate	*SE*	*Z*	*p*
Fixed effects				
Intercept	0.11	0.17	0.69	.49
Input Group = four verbs	0.23	0.33	0.68	.49
Verb Familiarity = unfamiliar	−0.06	0.17	−0.38	.70
Day = day three	0.10	0.17	0.58	.56
Input Group × Verb Familiarity	0.17	0.33	0.50	.62
Random effects				
		SD		
Participant (intercept)		0.92		

significant (A: r^2 = .01, p = .61; D: r^2 = .06, p = .31). Overall, the pattern suggests that the three experimental tasks are drawing on the same knowledge, but that cross-task correlations are more pronounced given four input verbs, possibly reflecting the stronger overall learning in that condition (if we assume that stronger learning leads to more systematic responses across tasks).

Experiment 1 Discussion

In Experiment 1, 5 year olds were exposed to a novel construction that mapped a novel form (VN_1N_2) to a novel semantics (an approach event—with N_2 as agent and N_1 as goal). The results demonstrate that 5-year-old children are able to glean some knowledge of this novel construction after quite minimal input. Even on day one, after exposure to just 16 input sentences, children recognized that VN_2N_1 utterances conveyed an approach event in which one of the Ns approaches the other (recall the high percentage of approach actions in the act out task); conversely, they were also able to recognize that approach actions should be described by a VNN structure (recall the high percentage of VNN productions in the production task). Performance improved with exposure (recall the effect of day

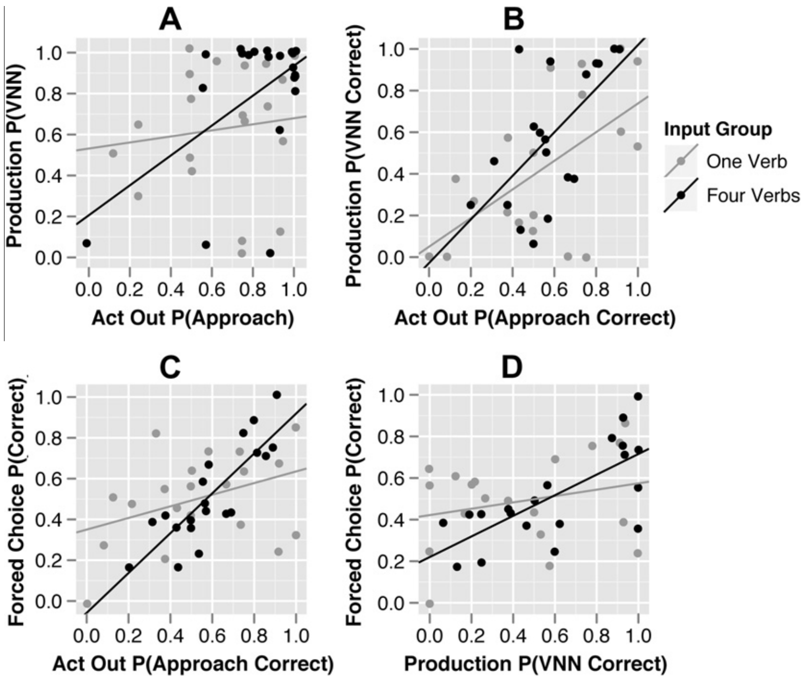

Fig. 5. Relationships among the three Experiment 1 tasks in the one verb (gray) and four verbs (black) conditions. The A plot depicts the relationship between the two measures of global construction learning. The B–D plots depict the relationship between various measures of linking rule acquisition.

in production and act out tasks), and importantly, as children improved on these global measures of construction learning, they also showed learning of the more precise linking of noun order to thematic role assignment (recall that the effect of day specifically held for VNN productions and approach actions with the correct *goal–agent* ordering, but not for those with the incorrect *agent–goal* ordering). In addition, we saw evidence that children's performance on the different tests was correlated, particularly for the four verbs group where performance was strongest, suggesting that the different tests tap the same developing knowledge of the mapping between the novel form and meaning.

A central question, however, is whether children are acquiring a construction that is tied to the particular verbs with which it has been encountered, or whether it can be generalized to unattested verbs. This was explored in the comparison between familiar and unfamiliar verbs, with the likelihood of VNN productions and approach act outs as the two dependent measures (i.e. the global measures of construction learning gleaned from the production and act out tests). The critical finding was that generalization was related to type frequency: children who had been exposed to input in which the construction was presented with four verbs showed good generalization (strong performance in the unfamiliar verbs condition), while those exposed to input in which the construction was presented with a single verb showed poor generalization (weaker performance in the unfamiliar verbs condition).

It is also interesting that the resistance to generalization when exposed to only a single novel verb holds steady on day three. Thus though children's overall knowledge of the mapping involved in the construction appears to be growing gradually more robust with increased exposure, the ability to generalize across verbs more directly relates to input structure. This indicates two different types of learning involved in forming construction generalizations, a point to which we return in the general discussion.

We also considered the possibility that construction learning could somehow be related to learning of the semantics of the verb(s) in the input, which is (unsurprisingly) stronger in the one verb condition. In fact, we only found evidence of a relationship between enacting the verb semantics and enacting the construction semantics for the unfamiliar novel verbs. These verbs were generally equally well enacted children in the four verb and one verb conditions, presumably because they were presented during testing rather than recalled from the input, however the detriment in construction semantics seen in the one verb condition compared to the four verbs condition (i.e. the type frequency effect) is only evident for trials in which the verb manner was correctly enacted. We suggest that in trials where verb manner was incorrect, children's performance was more generally haphazard, reflecting general uncertainty as to what to do in the task. Conversely, in trials where verb semantics was correct, their performance on construction semantics reflects what they have learned about that construction from their particular input. We also point out that the fact that we see different learning across the conditions supports our more general claim that children's performance in the act out and production tasks, particularly on the "global" measures of construction learning which we have primarily used, genuinely results from learning about the link between the construction's form and meaning. For example, if they were simply mimicking the types of actions/sentences they had witnessed in the input, it is not clear why they should specifically fail to perform in this way with unfamiliar verbs in the one verb condition.

The one task where we saw apparently weak performance across the board was the forced choice comprehension of linking. Recall that, unlike in production and act out comprehension, this test did *not* allow children to demonstrate what they had learned about the global semantics of the construction. Instead, every test trial asked children to choose between correct approach actions and foils which were identical except that the agent and goal participants were reversed. Children did not perform above "chance" (50%) in this task. However recall that we did see some evidence of learning of the specific *goal–agent* mapping in the act out and production tasks insofar as children were more likely to produce actions/sentences with correct linking on day three than day one. One possibility is that children may start the experiment with a preference for the alternative *agent–goal* word order. In that case, although at first blush performance in forced choice comprehension does not appear to show evidence that the ordering of semantic roles was learned, it is difficult to interpret the results without a baseline of how often VSO interpretation should be expected. What is needed is a control condition that allows us to see children's baseline preference for each of the two possible word orders.

Therefore in Experiment 2 we investigate whether children bring to the experiment a bias for agent-first word order, in order to provide a more appropriate control for Experiment 1. To that end we expose children to the *form* but not the semantics of the construction, so that they have no opportunity to learn the link between form and meaning. This allows us to additionally see whether children have any inherent bias to assume that the VNN form is linked to approach semantics.

Experiment 2 also allows us to address a potential problem in evaluating the results of the global analysis for the production test. Since we do not require children to produce the verb, one possibility is that their production of the two nouns may simply reflect an understanding that the pragmatics associated with having to describe a two-participant event, i.e. that they should probably name both event participants. This would not account for the difference between one verb and four verb conditions, since clearly the same pragmatics applies across conditions. Moreover, the fact that there are correlations between producing VNN responses in the production test, and approach actions in the act out comprehension test (particularly in the four verbs condition where learning is strongest), and that both increase with exposure, strongly suggests that these data are driven by learning. Nevertheless, it is of interest to test the extent to which producing a VNN sentence relies on exposure to the semantics of the novel construction.

The results of this experiment are compared both to chance and to the relevant results of Experiment 1.

Please cite this article in press as: Wonnacott, E., et al. Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language* (2012), doi:10.1016/j.jml.2011.11.004

Experiment 2: investigation of baseline performance

Methods

Participants

Thirty-five children (18 females and 17 males; age range 4:2–6:2; *M* = 5:2) were recruited from the same population as the Experiment 1 participants to take part in Experiment 2. All children were native speakers of English. As in Experiment 1, half were pseudo-randomly assigned to participate in the one verb group while the other half participated in the four verbs group.

Procedure

The procedure used in Experiment 2 was exactly the same as the Experiment 1 procedure, with two exceptions. First, in exposure trials, after the experimenter read a sentence and gave children the opportunity to repeat it, the sentence's meaning was *not* enacted. Children in Experiment 2 thus had exposure to the novel construction's form, but not its semantics. This was by design so that the input that children in Experiment 2 had was linguistic in nature, but did not strictly constitute exposure to a construction, since constructions are by definition form-function pairings. The Experiment 2 children thus provided a reasonable control to Experiment 1: if test performance in Experiment 2 ended up being worse than in Experiment 1, we can infer that only exposure to a consistent form-meaning pairing in Experiment 1 allowed for learning of the novel construction's characteristics.

Second, the procedure terminated after familiar verb testing on day one. Recall that at test children always completed the act out task, followed by the production task, followed by the forced choice task, and that in the forced choice task children viewed two approach movies paired with a VNN voiceover. This kind of content unfortunately constitutes input that might allow for learning of the novel construction's general meaning. Early termination of the procedure was intended to minimize this sort of exposure so that the children in Experiment 2 would be relatively untainted for use as controls for Experiment 1. Thus because testing in Experiment 2 covered only test items using familiar verbs on day one, neither verb familiarity nor day were manipulated.

All other aspects of the Experiment 2 materials and procedure were identical to Experiment 1.

Results

A recording error led to the loss of all responses from two children in the forced choice task. All forced choice analyses consequently reflect data from 33 rather than 35 participants.

Proportional condition means from Experiment 2 are summarized in Table 14 and Fig. 6. Detailed results for each task are discussed below.

Act out comprehension

Children were given this test having had no exposure to the semantics of the novel construction and so were essentially being asked to guess at a possible meaning. No child acted out an approach action in any of the trials: 100% of actions were assigned the other code according to the criteria outlined in Experiment 1. In fact, in 92% of these trials children performed no action at all. Note that because no approach actions were produced it was not possible to look at linking rules in the act out data.

Production

Experiment 2 production data were used to address two questions. The first is whether children's tendency to produce VNN utterances in Experiment 1 resulted from learning the form-meaning association exemplified by the novel construction, or whether this general behavior, if not the difference across conditions, might instead be explained in terms of the pragmatics of the experimental situation. The second question deals with linking rules. In order to determine whether the roughly 50% performance on the comprehension of linking trials in Experiment 1 was due to chance, we need to know what children do on linking trials if they have had no prior exposure to the linking between form and meaning.

Note that because children in Experiment 2 were tested on familiar verbs on day one only, the Experiment 1 data were made equivalent by excluding all trials involving unfamiliar verbs, and all trials occurring on day three. It follows from this that verb familiarity and day did not vary in the subsequent analyses.

Pragmatics and VNN productions. One concern in Experiment 1 was that children might be producing VNN sentences in response to the pragmatics involved in describing two-participant events. Recall also that the children in Experiment 2 witnessed VNN sentences during exposure (although the VNN sentences were given no interpretation). These same pragmatics apply in Experiment 2, which explains why—in the absence of knowing anything about the form-meaning association embodied by the novel construction—children still produced VNN responses 56% of the time. The important question, however, is whether children in Experiment 1 were more likely than children in Experiment 2 to produce VNN responses. If so, then children's behavior in Experiment 1 would indicate that the exposure in the first experiment resulted in true learning about the novel construction. To evaluate this possibility, a mixed model of the likelihood of a VNN response was fit to the Experiment 1 and Experiment 2 data. This included random participant intercepts and a single fixed effect for experiment. The results show that children in Experiment 1 were significantly more likely than children in Experiment 2 to produce a VNN response ($\beta = 2.18$, $z = 2.60$, $p = .0093$; Experiment 1: 77%; Experiment 2: 56%). The production of VNN sentences is thus not solely a response to situation pragmatics, nor to having previously heard similar such sentences. The link to semantics provided during exposure in Experiment 1 appears to be critically important.

Linking rules in Experiments 1 and 2. Was there input-dependent learning of linking rules in Experiment 1? Analysis of the Experiment 1 data showed improvement

Please cite this article in press as: Wonnacott, E., et al. Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language* (2012), doi:10.1016/j.jml.2011.11.004

Table 14
Proportion of response types (*SE*) by condition in Experiment 2's act out comprehension, production, and forced choice comprehension tasks.

	Act out			Production			Forced choice
	Approach correct	Approach incorrect	Other	VNN correct	VNN incorrect	Other	Correct
One verb familiar	0.00 (0.00)	0.00 (0.00)	1.00 (0.00)	0.13 (0.05)	0.42 (0.09)	0.46 (0.11)	0.33 (0.08)
Four verbs familiar	0.00 (0.00)	0.00 (0.00)	1.00 (0.00)	0.22 (0.07)	0.35 (0.09)	0.43 (0.10)	0.28 (0.06)

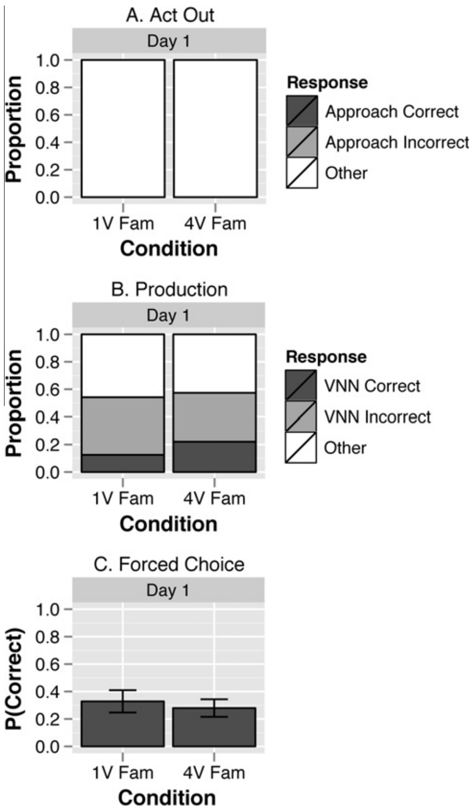

Fig. 6. Experiment 2: baseline condition results in the act out comprehension (A), production (B), and forced choice comprehension (C) tasks. Error bars in (C) show the standard error of the mean. For standard errors in (A) and (B), please refer to Table 14. 1*V* = one verb; 4*V* = four verbs; *Fam* = familiar.

in children's linking tendencies from day one to day three. Here we assess whether the acquisition of linking rules can additionally be established through comparison of Experiment 1 to the control provided by Experiment 2. Note that visual inspection of the day one familiar verb conditions in Fig. 2 suggests that when children in Experiment 1 produced a VNN response they used the correct (dark gray) and incorrect (light gray) linking rules at roughly equal rates. In contrast, the data in Fig. 6B suggests that when children in Experiment 2 produced a VNN response, they were biased towards use of the incorrect linking rules—i.e., there were more VNN incorrect (light gray) than VNN correct (dark gray) responses. This pattern in Experiment 2 is not surprising given other findings demonstrating that native speakers of English preferentially interpret the first noun of an NN sequence as the agent (Boyd, Gottschalk, & Goldberg, 2009; Wonnacott et al., 2008). Of interest is whether exposure in Experiment 1 allowed for significant unlearning of this agent-first bias in favor of the agent-last linking pattern exemplified by the novel construction.

We reasoned that two results would have to hold in the present analysis in order to positively establish tentative linking rule acquisition in Experiment 1. First, the likelihood of a VNN correct response would have to be significantly higher in Experiment 1 than Experiment 2. Second, the likelihood of a VNN incorrect response would either have to be lower in Experiment 1 than Experiment 2, or equivalent across experiments. These outcomes were tested for by fitting two models to the Experiment 1 and Experiment 2 data—one to VNN correct responses and one to VNN incorrect responses. Both models included random participant intercepts and a fixed effect for experiment; the model of VNN incorrect also included random item intercepts. The results demonstrate that children in Experiment 1 were in fact significantly more likely than children in Experiment 2 to produce a VNN correct response ($\beta = 2.08$, $z = 2.73$, $p = .0063$; Experiment 1: 39%; Experiment 2: 17%), and that the likelihood of a VNN incorrect response did not differ across experiments ($\beta = -0.03$, $z = -0.05$, $p = .96$; Experiment 1: 38%; Experiment 2: 39%). This indicates that the significant effect of experiment reported above in the comparison of Experiment 1 and 2 on VNN was specifically due to increased VNN correct responding in Experiment 1, and not simply to increased production of VNN responses, regardless of linking rules. Exposure to the novel construction's form and meaning in Experiment 1 thus led to significant learning about the construction's linking rules.

Forced choice comprehension

The forced choice task is a pure test of children's linking tendencies. In the absence of input indicating an alternative, native speakers of English should prefer to interpret the first noun of an NN sequence as the agent. This predicts that children in Experiment 2 should be biased towards choosing the wrong movie. The rate at which they pick the correct movie should therefore be below 50%. This prediction was tested in a mixed model of the likelihood of a correct response in the Experiment 2 data. The model included random participant intercepts and an estimate of the overall model intercept that was significantly below zero ($\beta = -1.02$, $z = -3.78$, $p = .00016$), which indicates that the likelihood of a correct response was indeed below 50%. This result verifies the existence of an agent-first linking bias in Experiment 2.

In order to determine whether exposure in Experiment 1 led to significant learning of the novel construction's linking rules, a model predicting the likelihood of a correct

Please cite this article in press as: Wonnacott, E., et al. Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language* (2012), doi:10.1016/j.jml.2011.11.004

response was fit to the Experiment 1 and 2 data. The model included a fixed effect of experiment and random participant intercepts. The results indicate that children in Experiment 1 were significantly more likely than children in Experiment 2 to pick the correct movie ($\beta = 1.22$, $z = 3.23$, $p = .0013$; Experiment 2: 53%; Experiment 1: 30%). Exposure to the novel construction in Experiment 1 thus led to linking rule learning that partially overcame the agent-first bias exemplified the Experiment 2.

Experiment 2 Discussion

The most important finding in Experiment 2 is that, in the absence of input that involves a distinct ordering, children have a clear preference for a *verb agent goal* ordering over a *verb goal agent* ordering. This was seen in the production task, and perhaps most clearly in the forced choice comprehension task, which focused specifically on linking. The comparison with Experiment 1 shows that, compared to the baseline provided by children in Experiment 2, there is clear evidence of learning of linking rule acquisition even on day one, after just 16 exposure trials. Note that the bias for the agent first ordering is in line with findings of other experiments with adult learners, (see e.g., Boyd et al., 2009; Wonnacott et al., 2008). The bias may stem from the fact that English generally orders agents before goals (Wonnacott et al. found a similar bias with a fully artificial language, but transfer may occur in that situation also), or it may be attributable to agents' greater conceptual accessibility—agents garner more visual attention because they are the entities in the events that move (see also Bever, 1970). We cannot tell at this point whether the bias is a result of transfer from English, or whether it is a universal bias. Regardless, it is now clear that the children who were exposed to the novel construction in Experiment 1 were able to overcome this bias to a significant extent as a result of learning. This finding is also in line with the day effects on linking that were documented in the Experiment 1 act out and production tasks: children move even further from baseline with additional exposure. Based on these results, we predict that sufficient exposure to the novel construction would eventually lead to linking rule performance at or near ceiling.

The results also confirm that the production of VNN sentences in Experiment 1 was not solely a response to the pragmatics of the testing situation. The same pragmatics apply in Experiment 2, and yet children produced significantly fewer VNN utterances. Moreover this condition also provided matched exposure to the VNN *form* – what was missing was the accompanying semantics – i.e. only in Experiment 1 are they actually exposed to the pairing of form *and* semantics: a novel construction. The different results strongly indicate that children's VNN responses in the production task in Experiment 1 were driven by construction learning (in line with the fact that production of such responses differed across conditions).

Finally, the results of the act out comprehension task demonstrate that children have no a priori bias to produce approach actions in the absence of input.

In summary, the results of Experiment 2 provide a relevant baseline against which to compare those of Experiment 1. The comparison indicates that children in Experiment 1 were indeed learning the construction, including the specifics of the linking pattern.

General discussion

The current work probes children's ability to acquire a novel verb argument structure construction under experimental conditions. We saw evidence, for the first time, that children as young as 5 years old were able to learn both the global semantics of a novel construction, and also the more precise relationship between the ordering of the NPs and semantic role assignment in an experimental context. Children's understanding of the construction was demonstrated in tests of both comprehension and production.

At the same time, children's generalizations appear to be tentative and are closely tied to the input received. First, in the production and act out comprehension tasks, we saw an effect of *quantity* of input on the robustness and precision of children's representations. At first test (day one), they already show a good understanding of the global mapping between *VNN* and *approach* event, producing VNN sentences in response to approach events 69% of the time, and approach events in response to VNN sentences 68% of the time. They also show some evidence of having learned the more precise link between noun order and thematic roles; for example, in production although only 48% of the VNN which the children produced expressed the goal as N_1 and the agent as $N_{2,}$ this is still significantly more than in the control provide by Experiment 2 where 31% of their VNN utterances had this mapping. On the other hand, this is rather far from the kind of deterministic behavior that is characteristic of adult-like knowledge of linking rules. Importantly, we do see a clear improvement in this type of behavior after more exposure (i.e. on day three) with children producing VNN sentences with the correct mapping 58% of the time, and approach events with the correct mapping 63% of the time. We suggest that performance would continue to increase with increased input, and that children's knowledge of the mappings involved in the construction are gradually solidifying with exposure. Note that this increase in performance was present for both old and new verbs.

Second, in addition to quantity of input, the *structure* of the input was also important. This effect was seen when we look at the learning of the global mapping between the VNN form and approach semantics in the production and act out comprehension tasks. Here there was a consistent interaction between the type of input children received and their ability to generalize the construction to unattested verbs. In both act out comprehension and production tasks, despite identical task demands, we found that children were markedly better at generalizing to a new verb if the construction had been exemplified with four verbs than if it had been exemplified with just one.

This difference held steady across day one and day three, suggesting that the difficulty in generalization in the one verb condition is not due to an insufficient quantity of input, but rather a response to the structure of the input. This raises the possibility that additional exposure to the 1verb condition would not improve performance on novel verbs (unless of course the input were changed so that the children went onto encounter the construction with more verbs). Thus in this data we see that it is possible to differentiate construction learning in the sense of generalization to new verbs (which seems here to depend on input structure) from improvement in knowledge of the mappings involved in the construction, be those tied to the verb on not, which gradually solidify with additional input. We feel that this is an important point that has been rather overlooked in the previous literature, which has tended to focus on the first type of generalization to new verbs (cf. also Johnson & Goldberg, submitted for publication b).

Returning to children's difficulty in generalizing on the basis of a single verb, it is important to note that, since we do not see the same difficulty in the four verb condition, this cannot be an overall problem with novelty, nor a general problem with producing or comprehending unfamiliar verbs. Rather it seems that encountering the construction with multiple verbs is important in allowing children to extend it to unfamiliar verbs. This effect of type frequency is familiar from work in theoretical linguistics (Barðdal, 2008; Bybee, 1985), and has been demonstrated in the case of morphology (e.g., Plunkett & Marchman, 1991) and in artificial grammar learning tasks (Childers & Tomasello, 2001; Gomez & Gerken, 1999; Suttle & Goldberg, 2011). Here we see experimental evidence for it for the first time with children, in both comprehension and production, in a novel construction learning task that involves both form and function.

One way of characterizing type frequency effects is in terms of rational or evidence-based learning, where the learner can entertain both generalized and lexically specific constructions, but determines constructional range on the basis of the sample of examples in the input. If there is no variability at the verb slot, so that every example of a construction occurs with a single verb, the evidence suggests that construction is not separable from the verb (if it were, it would be unlikely that there were no examples of that behavior in the sample). This would explain why children in the one verb condition are hesitant to demonstrate the construction with a new verb, instead falling back on the intransitive semantics with which that verb was modeled. This explanation is in line with a Bayesian perspective on learning, as adopted in other work with adults and children (Wonnacott, 2011; Gerken, 2006; Wonnacott et al., 2008; Xu & Tenenbaum, 2007; see also Perfors & Wonnacott, 2011). Note, however, that this type of explanation is not intended to be mechanistic; there is no claim that the child is explicitly making this type of inference. It is possible to view the problem in terms of what we take the task to be from the child's perspective: extracting mappings between form and function from the set of input sentences. To be able to generalize the construction across words, the child must form the abstract association: *Verb* N_1 N_2: N_2 *approaches* N_1. While it is possible in principle to extract this association from the one verb input, it might be easier for children to instead associate the semantics "N_2 *approaches*(*-in-manner*$_x$) N_1", with the less abstract *verb*$_i$ N_1 N_2. In this case, variability in verbs may be necessary for the child to be able to see that the construction can be disassociated from a particular verb and so create a general verb slot. More work is needed to probe these two types of explanation, and to determine the extent to which the second might function as an implementation of the first.

An interesting question is whether children and adults differ in this type of input driven generalization. Recall from the introduction that Boyd and Goldberg (2011) also found that 5-year-old children were conservative in their comprehension of a newly learned construction, failing to generalize fully to novel vocabulary. As discussed in the introduction, some aspects of the structure of the input in that experiment might also have favoured a more conservative interpretation of the new construction. In particular, although the construction was exemplified with multiple verbs, it always occurred with the same two nouns, possibly explaining why the children had particular difficulty generalizing when the nouns were also novel.[4] However this same study also found age differences: adults and 7 year old children did generalize; only the younger 5 year olds did not.

Given this previous finding, we have also run a pilot study on adults using the same materials and method as the current experiment. We found that adults generalized the construction fully to unfamiliar verbs in all tasks, even when exposed to the one verb input set. That is, adults did not show the same effect of low type frequency as children. However adults performed at ceiling on all measures in this pilot, making it precipitant to draw strong conclusions from this data. It is also possible that adults might actually form the same kinds of conservative representations from the input as children, but go beyond these in testing due to experimental pragmatics (i.e. they realize what the experimenter is asking of them, irrespective of whether that is warranted by the input). On the other hand, though we certainly would *not* wish to claim that adult generalization is never affected by type frequency (there is evidence elsewhere that it is, Barðdal, 2008; Suttle & Goldberg, 2011; Wonnacott et al., 2008), this result does at least suggest that adults generalize more readily than children given the same input.

Taken together with the age effects in Boyd and Goldberg (2011), it seems likely that input driven learning more generally interacts with different learners' biases at different ages, with 5 year olds biased to adopt more concrete representations than adults and older children. Note that it is possible to incorporate such an interaction with age into the input driven framework discussed above by assuming that children and adults come to the experiment with different biases. For example, adults may have learned from their previous exposure to English that word order constructions usually apply across lexical items; a Bayesian framework allows them to make use of this higher order knowledge about construction behavior so that it could outweigh even the very constrained input for this particular

[4] On the other hand, unpublished results that involved extremely variable input failed to result in generalization by younger children. There may well be a sweet spot, whereby input that is variable, but not *too* variable is ideal for forming a constructional generalization. This general idea is discussed below.

construction (Perfors & Wonnacott, 2011; Wonnacott et al., 2008). Older learners might also have an inherent bias towards more abstract representations (e.g., Rovee-Collier, 1999) (they might find it inherently easier to generalize from a single verb to a more abstract verb slot).

However these ideas about age related differences are currently speculative. It remains for future work to further probe the locus of age related differences, both to rule out an explanation in terms of experiment pragmatics, and to probe the specific type of input which do or do not lead to conservatism. A further question is whether younger children are even more conservative potentially accounting for the very high levels of conservatism reported for children of 3 years and under (e.g. Tomasello, 2000).

Another interesting question is exactly how precise children's construction representations are in terms of the way in which semantics is represented. We have assumed that children are gradually mapping the first and second N positions in VN_1N_2 respectively to the goal and agent roles in an approach event (whether verb-specifically or more generally), since there is good reason to think that their performance in the Act Out and Production tasks reflects learning of a general *VNN* to *approach mapping*. In particular, recall the fact that we see the same effect of input structure on the production of VNN sentences in response to approach actions and vice versa; also, performance on these two types of productions is correlated across tasks. We also see that participants become more likely to produce sentences/actions which reflect the agent–goal mapping with increased exposure. It is possible, however, that children may also have learned something about word order at a more general level. For example, they might have learned that "agents" come at the end of a *VNN* sentence. This could be tested using a forced choice task where children heard VNN sentences and choose between two non-approach scenes (e.g. one in which a giraffe moves away from a gorilla, one if which a gorilla moves away from a giraffe). Of particular interest would be whether this type of generalization would also be driven by the structure of the input (would there be more generalizations for more similar types of events? Would providing input sentences that were more/less variable along this dimension affect generalization?) These are interesting questions for future research.[5]

The picture that emerges from this work is in line with recent work that offers a nuanced account of construction learning (Abbot-Smith et al., 2008). The question is not whether children have an abstract representation, but rather how precise, robust and generalized their representations. Note, however, that given the age of our child participants, our data cannot speak to the contentious issue of whether children bring innate knowledge to the task of construction learning. For example, they are clearly able to recognize semantic roles such as agent and goal, and to link those roles to the sequential structure of sentences, but we cannot say that this is not due to their existing knowledge of English constructions. Nevertheless, we see clear evidence that a novel correspondence between word order and semantic roles can develop quite rapidly, albeit in a manner that is closely related to the quantity and type of input that is received.

[5] We thank an anonymous reviewer for pointing out this very interesting line of enquiry, and for suggesting the how it could be tested.

Conclusion

The present experiments demonstrate that children as young as 5 years old are capable of generalizing an abstract formal pattern and abstract meaning, beyond their input, based on minimal exposure. The novel construction used involved a cross-linguistically marked phrasal pattern that has the form, VN_1N_2, and the novel abstract meaning, "N_2 approaches N_1". When compared with controls, who heard only the VNN form of the construction without witnessing any corresponding semantics, 5-year-olds demonstrate significant generalization of the form and function of the construction to new utterances, as measured in both production and comprehension tasks. At the same time, closer inspection shows that children are keenly sensitive to the input: they learn the constructions more precisely and more robustly with more exposure, and generalize better to unfamiliar verbs if the input contains a variety of instances of that construction with different verbs (higher type frequency). Future work is needed in order to determine whether this strong input driven conservatism also occurs in older learners or is also a function of the children's age.

Acknowledgments

This research was supported by a British Academy Postdoctoral Research Fellowship and an ESRC grant (Award Reference: RES-062-23-0983) awarded to the first author, an NICHD training fellowship (2-T32-HD055272A) and a Kavli Innovative Research Grant (2011-009) awarded to the second author, and an NSF grant (# 0613227) and an Einstein fellowship from Berlin awarded to the final author. Thank you to Jacinth Greywoode for her help with data coding. Finally, we are very grateful to all the teachers, parents and children who made the research possible.

Appendix A. Novel verbs used in Experiments 1–3

Verb	Manner of approach
Chadding	Hopping on head
Biceing	Moving forwards while facing forwards
Datting	Moving forwards while facing backwards
Gaking	Sliding on stomach
Jibbing	Sliding forwards while lying on back
Ketting	Sliding on head
Migging	Tumbling
Nassing	Zig-zagging
Pemming	Sliding forwards while seated
Ralling	Spinning
Shuming	Sliding from side to side
Suzzing	Flying up and down
Teeping	Jumping
Wugging	Flying a level trajectory

Appendix B. Other responses in the Experiment 1 act out task

Test verb type	Input condition	% Moved both: N_1 and N_2 moved independently	% Transitive action: either N_1 acted onN_2 or N_2 acted on N_1	% Moved one: either N_1 or N_2 moved	% Moved neither: neither N_1 or N_2 moved
Old	Four verbs	6% (*SE* 3%)	0% (*SE* 0%)	1% (*SE* 3%)	8% (*SE* 4%)
New		14% (*SE* 6%)	0.6% (*SE* 0.6%)	3% (*SE* 2%)	4% (*SE* 3%)
Old	One verb	5% (*SE* 3%)	5% (*SE* 3%)	0% (*SE* 0%)	4% (*SE* 3%)
New		50% (*SE* 9%)	3% (*SE* 3%)	4% (*SE* 3%)	0% (*SE* 0%)

Appendix C. Other responses in the Experiment 1 production task

Test verb type	Input condition	% VN, mention goal v-goal	% VN, mention agent v-agent	% Other[a]	% Nothing neither N_1 or N_2 moved
Old	Four verbs	1% (*SE* 4%)	8% (*SE* 6%)	10% (*SE* 5%)	0% (*SE* 0%)
New		1% (*SE* 1%)	9% (*SE* 5%)	8% (*SE* 4%)	3% (*SE* 2%)
Old	One verb	6% (*SE* 5%)	8% (*SE* 6%)	8% (*SE* 3%)	0% (*SE* 0%)
New		10% (*SE* 5%)	18% (*SE* 7%)	13% (*SE* 6%)	13% (*SE* 6%)

[a] The category ‘‘other’’ includes trials in which the child did not produce the correct animal names, trials in which they produced other combination of the verb and nouns (e.g. ‘‘datting gorilla bear gorilla’’; ‘‘datting bear datting’’), and trials where they produced some English or pseudo-English sentence (e.g. ‘‘bear sliding’’; ‘‘rolling along bear’’).

References

Abbot-Smith, K., Lieven, E., & Tomasello, M. (2001). What preschool children do and do not do with ungrammatical word orders. *Cognitive Development, 16*, 679–692.

Abbot-Smith, K., Lieven, E. V. M., & Tomasello, M. (2008). Graded representations in the acquisition of English and German transitive constructions. *Cognitive Development, 23*, 48–66.

Ahrens, Kathleen (1995). The meaning of double object construction in Chinese. In *Proceedings of the sixth North American conference on Chinese linguistics* (Vol. 1, pp. 1–10). Los Angeles: GSIL, University of Southern California.

Akhtar, N. (1999). Acquiring word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language, 26*, 339–356.

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*, 952–965.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

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

Barðdal, Jóhanna. (2008). *Productivity: Evidence form case and argument structure in Icelandic*. Amsterdam: John Benjamins.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Bever, T. (1970). The cognitive basis for linguistic structures. In R. Hayes (Ed.). *Cognition and the development of language* (Vol. 279, pp. 279–362). New York: Wiley and Sons, Inc.

Bolinger, D. (1977). *Meaning and form*. London: Longman.

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di Semantica, 3*, 5–66.

Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *Journal of Modern Language, 93*, 418–429.

Boyd, J. K., & Goldberg, A. E. (2011). Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Language*.

Boyd, J. K., Gottschalk, E., & Goldberg, A. E. (2009). Linking rule acquisition in novel phrasal constructions. *Language Learning, 59*, 64–89.

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

Braine, M. D. S. (1976). *Children's first word combinations. Monographs of the society for research in child development*. Chicago: University of Chicago Press.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins Publishing Company.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping of a phrasal form and meaning. *Developmental Science, 8*, 500–508.

Chang, F., Kobayashi, T., & Amano, S. (2009). Social factors in the acquisition of a new word order. *First Language, 29*, 427–445.

Childers, J., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37*, 730–748.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, T. (2008). Young German children's early syntactic competence: A preferential looking study. *Developmental Science, 11*, 582–585.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Rinehart and Winston.

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

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

Gerken, L. A. (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*, 684–691.

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

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

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 14*, 289–316.

Goldwater, M. B., & Markman, A. B. (2009). Constructional sources of implicit agents in sentence comprehension. *Cognitive Linguistics, 20*, 675–702.

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

Gomez, R. L., & Gerken, L. D. A.-M. L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

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

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

Ingram, D., & Thompson, W. (1996). Early syntactic acquisition in German: Evidence for the modal hypothesis. *Language, 72*, 97–120.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Johnson, M., & Goldberg, A. E. (submitted for publication a). *Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs*. Princeton University.

Johnson, M., & Goldberg, A. E. (submitted for publication b) *Concrete similarity among exemplars facilitates the initial implicit learning of a constrained category*. Princeton University.

Kako, E. (2006). The semantics of syntactic frames. *Language and Cognitive Processes, 21*, 562–575.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory & Language, 43*, 508–529.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

Lieven, E., Pine, J., & Baldwin, G. (1997). Lexically-based learning and the development of grammar in early multi-word speech. *Journal of Child Language, 24*, 187–219.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. D. A.-J.-M. D. O.-D. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development, 20*, 121–136. doi:10.1016/JC.2004.08.001.

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

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

Perfors, A., & Wonnacott, E. (2011). Bayesian modeling of sources of constraint in language acquisition. In I. Arnon & E. Clark (Eds.), *Experience, variation, and generalization: Learning a first language* (pp. 277–294). John Benjamins Publishing Company.

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

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

Quené, Hugo., & van den Bergh, Huub (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413–425.

R Development Core Team (2010). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rovee-Collier, C. (1999). The development of infant memory. *Current Directions in Psychological Science, 8*, 80–85.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*, 323–348.

Suttle, L., & Goldberg, A. E. (2011). Partial productivity of argument structure constructions. *Linguistics, 49*, 1237–1269.

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.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65*, 1–14.

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

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.

Year, J., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *Modern Language Journal, 93*, 399–417.