



Semantics versus statistics in the retreat from locative overgeneralization errors

Ben Ambridge*, Julian M. Pine, Caroline F. Rowland

University of Liverpool, United Kingdom

ARTICLE INFO

Article history:

Received 25 February 2011

Revised 7 January 2012

Accepted 13 January 2012

Available online 9 February 2012

Keywords:

Child language acquisition

Retreat from overgeneralization

No negative evidence problem

Entrenchment

Pre-emption

Semantic verb class

Verb semantics

Fit account

Grammaticality judgment

ABSTRACT

The present study investigated how children learn that some verbs may appear in the figure-locative but not the ground-locative construction (e.g., *Lisa poured water into the cup*; **Lisa poured the cup with water*), with some showing the opposite pattern (e.g., **Bart filled water into the cup*; *Bart filled the cup with water*), and others appearing in both (*Lisa sprayed water onto the flowers*; *Lisa sprayed the flowers with water*). Grammatical acceptability judgments were obtained for the use of each of 142 locative verbs (60 for children) in each sentence type. Overall, and for each age group individually, the judgment data were best explained by a model that included ratings of the extent to which each verb exhibits both the broad- and narrow-range semantic properties of the figure- and ground-locative constructions (relating mainly to *manner* and *end-state* respectively; Pinker, 1989) and the statistical-learning measure of overall verb frequency (entrenchment; Braine & Brooks, 1995). A second statistical-learning measure, frequency in each of the two locative constructions (pre-emption; Goldberg, 1995), was found to have no additional dissociable effect. We conclude by drawing together various theoretical proposals to arrive at a possible account of how semantics and statistics interact in the retreat from overgeneralization.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

It has been long been recognized that the defining characteristic of human language is speakers' ability to produce novel utterances, (i.e., utterances that have never been encountered in precisely that form) (e.g., Chomsky, 1957). How children acquire this productivity is therefore a question that lies at the very heart of language acquisition research.

A complicating factor is that many of the potential generalizations suggested by the data to which learners are exposed yield utterances that are deemed ungrammatical by adult native speakers. For example, many verbs that may appear in the *figure-locative* (or *contents-locative*) construction (e.g., *Lisa sprayed water onto the flowers*) may also

appear in the *ground-locative* (or *container-locative*) construction (e.g., *Lisa sprayed the flowers with water*). However, a speaker who formed the generalization that *all* verbs attested in the former construction may also appear in the latter would produce overgeneralization errors for verbs such as *pour* (e.g., *Lisa poured water into the cup*; **Lisa poured the cup with water*). On the other hand, a speaker who was unwilling to generalize *any* verbs into unattested constructions would not acquire adult-like productivity. Speakers must somehow arrive at exactly the right generalization: Certain verbs attested in only one of the two constructions generalize to the other construction, but many do not.

The question of how speakers acquire this partial, restricted productivity turns out to be an extremely difficult one to answer (see Bowerman, 1988). One possible answer is that children are conservative and avoid extending verbs into unattested constructions (e.g., Baker, 1979; Berwick & Weinberg, 1984). In fact, overgeneralization errors such as **Lisa filled water into the cup* are attested in both naturalistic

* Corresponding author. Address: School of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford St. South, Liverpool L69 7ZA, United Kingdom. Tel.: +44 151 794 1111.

E-mail address: Ben.Ambridge@Liverpool.ac.uk (B. Ambridge).

and experimental studies (e.g., Pinker, 1989). In any case, even if children made no overgeneralization errors (as may be the case for some individuals), we would still require an explanation of how children avoid these errors whilst retaining the capacity for productivity.

Another possible solution is that children retreat from error using corrective feedback from adults (e.g., a caregiver might respond to an overgeneralization such as **Lisa poured the cup with water with That's right, she poured water into the cup*). However, whilst such feedback no doubt aids in the retreat from overgeneralization (e.g., Chouinard & Clark, 2003), it is implausible to assume that every potential overgeneralization error was produced in childhood, with corrective feedback subsequently received. Indeed, a number of studies have shown that adults and children reject as ungrammatical particular sentences with novel verbs, for which they cannot have received such feedback (e.g., Ambridge, Pine, & Rowland, 2011; Ambridge, Pine, Rowland, & Young, 2008; Wonnacott, Newport, & Tanenhaus, 2008).

The aim of the present study is to evaluate two more successful proposals: Pinker's (1989) *semantic verb class hypothesis* and a *statistical learning* account (e.g., Boyd & Goldberg, 2011; Braine & Brooks, 1995; Goldberg, 1995; Stefanowitsch, 2008; Wonnacott et al., 2008). Although there exists some evidence in support of both proposals, the present study is the first to attempt to evaluate the relative contributions of each in a comprehensive manner (i.e., by including all the verbs listed as relevant to the chosen constructions). This study focuses on the locative alternation (e.g., *Lisa sprayed water onto the flowers/Lisa sprayed the flowers with water*) which, for reasons discussed below, constitutes a particularly stringent test case for each account. After outlining the semantic verb class and statistical learning hypotheses, we present the findings of a grammaticality judgment study that aims to tease apart the two accounts. We conclude by outlining a proposal for a hybrid learning mechanism that yields all the experimental effects observed.

1.1. The semantic verb class hypothesis

Pinker's (1989) semantic verb class hypothesis proposes a two-stage learning process. First, the child sets up a broad-range rule that, in effect, allows any verb that has appeared in a particular construction (here, the figure-locative) to also appear in a related construction (here, the ground-locative). For example, the broad-range rule for the locative would allow a child who heard a figure-locative sentence such as *Lisa loaded hay onto the wagon* to produce a ground-locative sentence such as *Lisa loaded the wagon with hay* (or vice versa). Note that this rule also generates overgeneralization errors for non-alternating verbs (e.g., *Lisa poured water into the cup* → **Lisa poured the cup with water*). An important, but often overlooked, aspect of Pinker's (1989) proposal is that the broad-range rule does not transform one syntactic structure into another *per se*. In fact, the rule operates at the level of verb semantics (Pinker, 1989, p. 79):

Basically, it is a gestalt shift: One can interpret *loading* as moving a theme (e.g., hay) to a location (e.g., a

wagon), but one can also interpret the same act in terms of changing the state of a theme (the wagon), in this case from empty to full, by means of moving something (the hay) into it.

Innately-specified semantics–syntax linking rules then spell out the argument structure for the intended sense of the verb (Pinker, 1989, p. 79, examples added):

In the old verb, the moving thing was the theme, and hence was linked to direct object [*Lisa loaded the hay. . .*]; in the new verb, the location is the theme...and hence is linked to object [*Lisa loaded the wagon. . .*].

Evidence for the existence of this broad-range rule comes from the holism requirement (e.g., Rappaport Hovav & Levin, 1985). The ground-object locative (e.g., *Lisa loaded the wagon with hay; Lisa taught the students French*) is felicitous only if the underlying object actually undergoes a state change (e.g., the wagon ends up full; the students end up knowing French). If this does not happen, the figure-object locative is much more felicitous (e.g., *Lisa loaded hay onto the wagon; Lisa taught French to the students*).

Parallel broad-range rules exist for the dative and causative alternations. For the dative (Pinker, 1989, p. 82), the rule takes a verb with the meaning component "X causes Y to go to Z" and generates a verb with the meaning component "X causes Z to have Y" (e.g., *Lisa sent the letter to Bart* → *Lisa sent Bart the letter*). In the same way as for the locative, if this semantic criterion (here, caused possession) is not met, the rule may not operate (e.g. one cannot say **Lisa sent Chicago the letter* as Chicago cannot possess the letter). For the causative (p. 88) the broad-range rule takes a verb with the meaning component "X goes to a location or state" and generates a verb with the meaning component "Y acts on X, causing X to go to a location or state" (e.g., *The ball rolled* → *Lisa rolled the ball*). Again, this semantic criterion (here, Y acting directly on X) must be met for the resulting utterance to be felicitous (e.g., one cannot say *Lisa crashed the car* if she simply distracted the driver, causing him to crash the car).

The broad-range rule constitutes a necessary, but not sufficient, criterion for undergoing the alternation. Under the second stage of the learning process, the child sets up narrow-range semantic verb classes: classes of verbs with similar meanings that have been attested in (a) the figure-locative construction only (e.g., *pour, spill, drip, dribble*), (b) the ground-locative construction only (e.g., *cover, fill, coat, line*), or (c) both constructions (e.g., *spray, splash, sprinkle, spatter*). After this point, a verb will be used in a particular construction only if either the verb itself or one of its semantic classmates has been attested in that construction. Any newly-encountered verb will be assimilated into an existing semantic class, taking on the argument-structure privileges of its classmates. Overgeneralization errors (formed using the broad-range rule) occur early in development as children have still to hone the narrow-range classes, and cease when the former is abandoned in favor of the latter, "presumably around puberty" (Pinker, 1989, p. 349).

Importantly, the clustering of verbs into figure-only, ground-only or alternating classes is not arbitrary. The

behavior of each class is determined by its compatibility with the broad-range rule. Figure-only (content-only) classes specify “the manner of (causation of) motion of a substance to a medium or container” (p. 77), because the figure-locative construes the substance (figure) as the affected entity. Ground-only (container-only) classes specify “that a surface, container, or medium undergoes a particular change resulting from the addition of material to it” (p. 77), because the ground-locative construes the container (ground) as the affect entity. Alternating classes specify both the manner and the endstate. The 15 narrow-range semantic classes that Pinker (1989, pp. 126–127) proposes for the locative alternation are listed in Table 1.

The semantic verb class hypothesis is sometimes characterized in the literature as an “all-or-nothing” account: a verb is either in a class that licenses a particular construction or it is not. For example, one anonymous reviewer of this paper interpreted Pinker’s (1989) account as claiming that “the primary outcome of learning is a deterministic system where any verb is either grammatical or ungrammatical in a particular construction. (Every verb is either in an alternating or in a non-alternating class)”. On this reading, the semantic verb class hypothesis struggles to explain gradient effects in grammaticality judgments (as we shall see shortly, some overgeneralization errors are rated as more unacceptable than others). However, an alternative, more probabilistic interpretation of Pinker’s (1989) account is also possible. As discussed above, classes vary with regard to the extent to which they are compatible with the semantic core of each construction. Thus, there is room within the theory for (for example) one figure-only class to be “more figure-only” than another, if the former is less compatible with the notion of state-change (the semantic core of the ground-locative

construction) than the latter. Indeed, for the causative alternation, Pinker (1989) presents naturalistic data suggesting that causativization errors are more common for intransitive-only classes that are nevertheless somewhat compatible with the notion of direct causation (e.g., *come*, *go*) than for classes that are entirely incompatible with this notion (e.g., *sing*, *talk*).

Another issue where there is room for debate is whether or not the semantic verb class hypothesis can account for verb frequency effects (as discussed below, overgeneralization errors are generally rated as more unacceptable for high frequency verbs than semantically-matched lower frequency verbs). On the one hand, a strictly deterministic view of semantic classes would seem to rule out any within-class frequency effects. On the other hand, it must be borne in mind that the classes are assumed to be formed on the basis of exposure to the input. Thus, whilst Pinker (1989) does not discuss this possibility specifically, it would be plausible to assume that lower frequency verbs take longer to be assigned to the appropriate semantic class, and hence are more susceptible to error.

Most evidence for the semantic verb class hypothesis comes from studies of the transitive–causative alternation. For example, in the study of Brooks and Tomasello (1999), children were taught two novel verbs; one semantically-consistent with a class of verbs that may appear in both the intransitive and transitive construction (similar in meaning to *spin/bounce*); the other semantically-consistent with an intransitive-only class (similar in meaning to *ascend/go up*). Crucially, however, both verbs were presented in intransitive constructions only during training. At test, the experimenter attempted to elicit transitive–causative utterances (e.g., *The mouse meeked the cup*). As predicted by the semantic verb class hypothesis, children aged 4;5

Table 1

Pinker’s (1989) Narrow-range semantic verb classes for the locative constructions. Note that class names (e.g., “Smear-type”) have been added by the present authors.

| |
|--|
| Figure- (content-) oriented (<i>into/onto verbs</i>) |
| Smear-type, Alternating (N = 10), designated reference category. Simultaneous, forceful contact and motion of a mass against a surface (<i>brush, dab, daub, plaster, rub, slather, smear, smudge, spread, streak</i>) |
| Stack-type, Alternating (N = 3). Vertical arrangement on a horizontal surface (<i>heap, pile, stack</i>) |
| Spray-type, Alternating (N = 7). Force is imparted to a mass, causing ballistic motion in a specified spatial direction along a trajectory (<i>inject, spatter, splash, splatter, spray, sprinkle, squirt</i>) |
| Scatter-type, Alternating (N = 4). Mass is caused to move in a widespread or non-directed distribution (<i>bestrew, scatter, sow, strew</i>) |
| Pour-type, Content-only (N = 10). A mass is enabled to move via the force of gravity (<i>dribble, drip, drizzle, dump, ladle, pour, shake, slop, slosh, spill</i>) |
| Coil-type, Content-only (N = 6). Flexible object extended in one dimension is put around another object (preposition is <i>around</i>) (<i>coil, spin, twirl, twist, whirl, wind</i>) |
| Spew-type, Content-only (N = 8). Mass is expelled from inside an entity (<i>emit, excrete, expectorate, expel, exude, secrete, spew, vomit</i>) |
| Glue-type, Content-only (N = 9). Verbs of attachment (<i>attach, fasten, glue, nail, paste, pin, staple, stick, tape</i>) |
| Ground- (container-) oriented (<i>with verbs</i>) |
| Stuff-type, Alternating (N = 6). A mass is forced into a container against the limits of its capacity (<i>pack, cram, crowd, jam, stuff, wad</i>) |
| Load-type, Container-only (N = 3). A mass of a size, shape, or type defined by the intended use of a container... is put into the container, enabling it to accomplish its function (<i>load, pack, stock</i>) |
| Cover-type, Container-only (N = 19). A layer completely covers a surface (<i>deluge, douse, flood, inundate, bandage, blanket, coat, cover, encrust, face, inlay, pad, pave, plate, shroud, smother, tile, fill, occupy</i>) |
| Pollute-type, Container-only (N = 22). Addition of an object or mass to a location causes an aesthetic or qualitative, often evaluative, change in the location (<i>adorn, burden, clutter, deck, dirty, embellish, emblazon, endow, enrich, festoon, garnish, imbue, infect, litter, ornament, pollute, replenish, season, soil, stain, tint, trim</i>) |
| Soak-type, Container-only (N = 15). A mass is caused to be coextensive with a solid or layer-like medium (<i>interlace, interlard, interleave, intersperse, interweave, lard, ripple, vein, drench, impregnate, infuse, saturate, soak, stain, suffuse</i>) |
| Clog-type, Container-only (N = 12). An object or mass impedes the free movement of, from, or through the object in which it is put (<i>block, choke, clog, dam, plug, stop up, bind, chain, entangle, lash, lasso, rope</i>) |
| Bombard-type, Container-only (N = 8). A set of objects is distributed over a surface (<i>bombard, blot, dapple, riddle, speckle, splotch, spot, stud</i>) |

and older (though not a younger group aged 2;5) produced fewer such utterances for the novel *ascending* verb than the novel *spinning* verb.

Similar results were observed in the grammaticality judgment studies of Ambridge et al. (2008, 2011). For example, when taught a novel verb meaning to laugh in a particular manner, children rated transitive–causative utterances (e.g., **The funny clown tammed Lisa*) as significantly less acceptable than intransitive utterances (e.g., *Lisa tammed*), as laughing verbs belong to an intransitive-only semantic class (e.g., **The funny clown laughed/giggled Lisa*). However, when taught a novel manner-of-motion verb, both uses (e.g., *Bart meeked the ball*; *The ball meeked*) were rated as equally acceptable, as manner-of-motion verbs form an alternating semantic class (e.g., *Bart rolled/bounced the ball*). Note that these studies provide evidence for the existence of the narrow-range classes, but do not address the question of whether or not the broad-range rule is psychologically real.

Conversely, the only study to investigate Pinker's (1989) hypothesis with regard to the locative constructions looked solely at the broad-range rule. Gropen, Holander, and Goldberg (1991a, 1991b) showed that children were more likely to use a ground-locative construction when a novel verb denoted a novel state-change action than a novel manner of motion. Thus, no locative study to date has investigated the role of the narrow-range classes, which are crucial in children's retreat from – or avoidance of – overgeneralization errors.

This is a significant omission because, compared to the causative alternation, the semantic classes proposed for the locative alternation would seem particularly difficult to acquire. For the former, the distinction between intransitive-only classes (e.g., *laugh, giggle, chuckle*) and alternating classes (e.g., *slide, roll, bounce*) is relatively clear-cut (internally- versus externally-caused events) and, importantly for children, amenable to direct observation. For the locative constructions, all that distinguishes some figure-only classes (e.g., *pour, spill*) from alternating classes (e.g., *spray, splash*) is whether the liquid is *caused* to move (as in the latter case) or merely *enabled* to move via the force of gravity (as in the former). Perhaps even more esoterically, all that distinguishes some ground-only classes (e.g., *fill, cover*) from alternating classes (e.g., *spray, splash*) is whether or not the receptacle ends up completely full or covered.

Thus, the first aim of the present study is to investigate the psychological reality of both the broad-range rule and the narrow-range semantic classes proposed for the locative constructions, by obtaining ratings for all 142 locative verbs listed by Pinker (1989).

1.2. Statistical learning accounts

The entrenchment hypothesis (Braine & Brooks, 1995) states that repeated presentation of a verb (e.g., *pour*) in one (or more) attested construction (e.g., the figure-locative; *Lisa poured water into the cup*) causes the learner to gradually form a probabilistic inference that adult speakers do not use that particular verb in non-attested constructions (e.g., the ground-locative **Lisa poured the cup with*

water). This hypothesis predicts that both (a) the rate of production and (b) the rated acceptability of overgeneralization errors will decrease with increasing verb frequency, which strengthens the inference that non-attested uses are ungrammatical.

Pre-emption (e.g., Goldberg, 1995) is similar, except that overgeneralization errors are probabilistically blocked not by any use of the relevant verb, but by a use of the verb that expresses the same intended meaning as the non-observed use. Thus, for example, ground-locative uses of *pour* (e.g., **Lisa poured the cup with water*) are probabilistically blocked by figure-locative uses of *pour* (e.g., *Lisa poured water into the cup*) but not by intransitive or dative uses (e.g., *The rain was pouring down*; *Lisa poured Homer a drink*). This account involves the learner computing a mismatch between observed and expected uses (e.g., noticing that a speaker used a figure-locative construction where a ground-locative construction might otherwise have been expected) and thus inferring that the latter is ungrammatical for that particular verb.

Both entrenchment and pre-emption enjoy support from production and judgment studies investigating transitive–causative overgeneralizations such as **The funny clown laughed/giggled Bart* (e.g., Ambridge et al., 2008; Brooks, Tomasello, Dodson, & Lewis, 1999; Ambridge, Pine, Rowland, Jones, & Clark, 2009; Ambridge & Lieven, 2011) and dative overgeneralizations such as **Bart said Lisa something funny* (Stefanowitsch, 2008, 2011; Ambridge, Pine, Rowland, & Chang, in press; Goldberg, 2011). However, to our knowledge, neither has been tested on the locative constructions, with the exception of a single sentence pair, which produced inconclusive results (Theakston, 2004). This omission is significant because, again, the locative alternation would seem to constitute a particularly stringent test case. Compared to the transitive–causative alternation, the relevant verbs (and presumably the locative constructions themselves, though we are not aware of any counts) are of considerably lower frequency, meaning that there has been considerably less opportunity for these statistical-learning processes to occur.

The second goal of the present study, then, is to test the entrenchment and pre-emption hypotheses for the figure- and ground-locative constructions. In doing so, we investigate two key questions that relate in particular to the relationship between entrenchment/pre-emption and verb semantics. The first is whether verb frequency effects are observed above and beyond effects of semantic verb class, or whether apparent frequency effects are semantic effects “in disguise”. Under this latter possibility, frequency effects arise simply because low frequency verbs take longer to be assigned to the appropriate semantic class than higher frequency verbs. The second question is whether, conversely, statistical learning can obviate the need for a learning process that is sensitive to verb semantics. For example, Stefanowitsch (2008, p. 527) suggests that “speakers might uncover certain semantic motivations for these constraints (for example, the ‘narrow-class rules’ suggested in some lexicalist approaches, e.g., Pinker, 1989), but those semantic motivations are not necessary for learning the constraint in the first place”.

1.3. The present study

The aim of the present study was to ascertain the relative contributions of Pinker's (1989) broad-range rule and narrow-range classes, entrenchment, and pre-emption to children's retreat from locative overgeneralization errors. To that end, ratings of figure- and ground-locative sentences for all 142 locative verbs listed by Pinker (1989) were obtained from adults. Children aged 5–6 and 9–10 rated a subset of 60 verbs (20 each of the figure-only, ground-only, and alternating type). We then investigated the ability of four measures to predict the pattern of judgments: (a) ratings of the extent to which each verb is consistent with the broad-range semantic core of the figure- and ground-locative constructions (the *manner* of an action versus its *end-state*), (b) ratings of the extent to which each verb exhibits the semantic properties that characterize each of Pinker's 15 narrow-range classes, (c) entrenchment (raw verb frequency) and (d) pre-emption (frequency of the verb in each locative construction).

2. Method

2.1. Participants

Participants were 48 children aged 5;10–6;8 ($M = 6;3$), 48 aged 9;10–10;9 ($M = 9;4$) and 30 adults (undergraduate students aged 18–21). Fewer adults than children were required because, unlike children, every adult completed every test trial. A further 30 adults were recruited as semantic raters. All participants were monolingual English speakers.

2.2. Design, materials and procedure

2.2.1. Grammaticality judgments

All 142 locative verbs listed by Pinker (1989) were included in the study. These comprised 33 figure-locative-only verbs, 76 ground-locative-only verbs and 33 alternating verbs. These classifications were checked against those given by Levin (1993) and found to match, with the exception of the single verb *bestrew* (alternating for Pinker, figure-only for Levin), and 12 verbs not listed by Levin (*streak*, *drizzle*, *dump*, *ladle*, *shake*, *wad*, *occupy*, *burden*, *infuse*, *chain*, *lasso*, *rope*).

2.2.1.1. Adults. Each adult rated the acceptability of one figure-locative and one ground-locative sentence for each verb, using a 7-point Likert scale. Both variants of each sentence used the same lexical items, except for the preposition *into/onto* which appeared in the figure-locative variant only. For example, the figure- and ground-locative sentences for *splash* for the first counterbalance group were *Marge splashed juice onto the carpet* and *Marge splashed the carpet with juice*. Adults completed a written questionnaire, of which there were six different versions; this comprised three different sets of sentences – each with a different combination of AGENT, FIGURE and GROUND – each presented in two different pseudo-random orders (with the stipulation that the same verb was

never presented in consecutive trials). The use of three different sets of sentences allows us to be confident that ratings reflect the acceptability of a particular verb in a particular locative construction, as opposed to the acceptability of one particular sentence (adding the factor of set did not improve the predictive ability of any of the regression analyses outlined below).

2.2.1.2. Children. Children rated a subset of 60 verbs; 20 of each type (for comparative purposes, some analyses of adult data were restricted to this “core set”, as opposed to the full “extended set”):

Figure (contents) only: *Dribble, Drip, Drizzle, Dump, Pour, Spill, Coil, Twirl, Whirl, Spew, Vomit, Attach, Fasten, Glue, Nail, Paste, Pin, Staple, Stick, Tape.*

Alternating: *Brush, Dab, Rub, Smudge, Spread, Heap, Pile, Stack, Splash, Splatter, Spray, Sprinkle, Squirt, Scatter, Cram, Crowd, Jam, Load, Pack, Stock.*

Ground (container) only: *Flood, Bandage, Coat, Cover, Fill, Clutter, Dirty, Infect, Litter, Stain, Ripple, Drench, Saturate, Soak, Stain, Block, Clog, Bind, Entangle, Splotch.*

Care was taken to ensure that each of the 15 narrow-range classes outlined by Pinker (1989; see Table 1) was represented in this core set. Because we felt children would be unable to complete 120 trials, verbs were split into three blocks, with each child completing only one (40 trials). Each block of trials included one figure- and one ground-locative sentence for each of 20 verbs. For children, sentences were presented auditorily, with accompanying cartoon animations.

Children provided their ratings using a 5-point “smiley face” scale (see Fig. 1), having first completed a training procedure designed to familiarize them with the correct use of the scale. Full details of the training and rating procedure can be found in Ambridge et al. (2008) and Ambridge (2011). In brief, children first select either a green or a red counter to indicate “broadly grammatical” or “broadly ungrammatical” respectively, before placing the counter on the scale to indicate the degree of (un)grammaticality. Training sentences (see Appendix A) – two fully grammatical, two fully ungrammatical and three intermediate – were used to illustrate the appropriate use of the scale, with the experimenter providing feedback where necessary.

After training, children completed the test trials in pseudo-random order, with the stipulation that the same verb was never presented in consecutive trials.

For all participants, the dependent measure for the main analysis was a “difference score” calculated by subtracting the rating for each ground-locative sentence (e.g., *Bart poured the cup with water*; *Lisa filled the cup with water*) from the rating for its equivalent figure-locative sentence (e.g., *Bart poured water into the cup*; **Lisa filled water into the cup*), regardless of which form was grammatical for the verb in question. Difference scores were calculated in this uniform way (rather than by subtracting the rating for each ungrammatical sentence from its grammatical equivalent) in order to allow for analyses that collapse across ground-only, figure-only and alternating verbs.

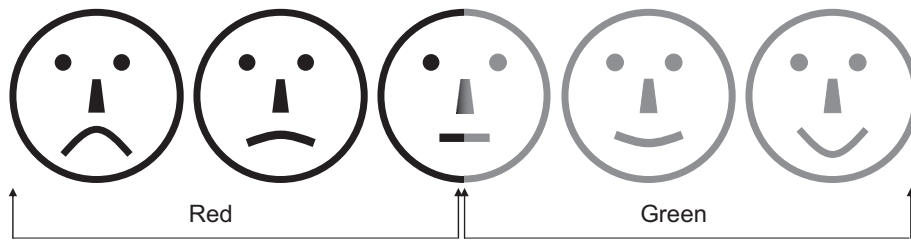


Fig. 1. The five-point rating scale used by children to rate sentences for grammatical acceptability (reproduced with permission from Ambridge et al. (2008)).

Difference scores are commonly used in grammaticality judgment studies of this nature (e.g., Pinker et al., 1987; Ambridge et al., 2008), in order to control for both the fact that participants generally prefer sentences with higher frequency verbs, and for any infelicities associated with particular test sentences. For example, if participants consider the chosen NPs to be less than fully appropriate for a given verb, this will presumably affect both locative variants equally, meaning that the difference score is unaffected. An additional analysis that constitutes a particularly stringent test of the pre-emption and entrenchment hypotheses was conducted on the raw rating data.

2.2.2. *Semantic (broad-range-rule) judgments*

In order to evaluate the contribution of Pinker's (1989) broad-range rule for the locative, it was necessary to obtain a measure of the extent to which each verb instantiates the semantic core of each construction. For each verb, 15 adult participants (who did not complete the grammaticality judgment task) were asked to rate, on a 10-point scale, the extent to which

- (a) The word describes the **particular manner/way** in which the action occurs (i.e., if the manner/way changes, this word is no longer an appropriate description of the event).
- (b) The word describes the **end-state** of an action (i.e. if the relevant end-state is not arrived at, this word is no longer an appropriate description of the event).

These definitions correspond to the semantic core of the figure (contents) and ground (container) locative respectively. Participants were given training examples (see Appendix B) that did not include the test verbs, but used simple intransitive and transitive verbs and manners/end-states not relevant to the locative constructions (e.g., quickly versus slowly, dead versus alive). This was to avoid alerting participants to the correlation between *manner* verbs and the figure-locative construction and *end-state* verbs and the ground-locative construction. Had participants been aware of this correlation, they may have completed the semantic rating task by generating both a figure- and a ground-locative sentence for each verb (e.g., *I poured water into the cup, *I poured the cup with water*),

Table 2
Original semantic features rated (third column) and the five composite semantic features extracted by Principal Components Analysis (right five columns).

| Class | Type | Original semantic feature rated by participants/Extracted Factors | Splattering (Alt/Fig) | Joining (Fig/Grd) | Stacking (Alt) | Gluing (Fig/Alt) | Smearing (Fig) |
|--------------------------|------|---|-----------------------|-------------------|----------------|------------------|----------------|
| Fig1 | Alt | Someone presses A against B and moves A around | .183 | .214 | -.321 | -.079 | .406 |
| Fig2 | Alt | Someone vertically arranges A on B | -.062 | .218 | .508 | -.153 | -.280 |
| Fig3 | Alt | Someone applies force to A causing it to "shoot off" onto B | .628 | -.135 | -.052 | .472 | -.001 |
| Fig4 | Alt | Someone causes A to move onto B in a widespread or non-directed distribution | .871 | .205 | .109 | -.103 | .102 |
| Fig5 | Fig | Someone enables A to move onto B via the force of gravity | .795 | .005 | .187 | .249 | -.070 |
| Fig6 | Fig | Someone puts A (a flexible object) around B | -.164 | .669 | -.182 | -.039 | .394 |
| Fig7 | Fig | Someone causes A to move from inside an entity, out onto B | .833 | -.017 | .168 | .338 | -.082 |
| Fig8a | Fig | Someone causes A to become attached to B | -.504 | .627 | -.143 | .406 | -.028 |
| Fig8b | Fig | The verb specifies some particular type of substance | .283 | .283 | -.108 | .624 | .039 |
| Grd1 | Alt | Someone forces A into B against the limits of B's capacity | -.101 | -.040 | .836 | -.092 | .293 |
| Grd2 | Alt | Someone puts A into B, allowing B to accomplish its function | -.352 | .078 | .394 | .200 | -. 510 |
| Grd3 | Grd | Someone causes A to completely cover B | .370 | .634 | .322 | -.398 | .073 |
| Grd4 | Grd | Someone adds A onto B causing an aesthetic change in B | .273 | .611 | -.272 | -.369 | -.380 |
| Grd5 | Grd | Someone causes A and B to become coextensive (i.e., "the same" object) | -.481 | .610 | -.063 | .215 | -.281 |
| Grd6 | Grd | Someone causes A to prevent (1) movement of B (2) movement of an object from B to another location or (3) movement of an object through B | -.503 | .324 | .464 | .323 | .375 |
| Grd7 | Grd | Someone distributes A (a set of objects) over B | .792 | .397 | .062 | -.107 | -.045 |
| <i>Extracted Factors</i> | | | | | | | |
| | | Eigenvalue | 4.34 | 2.49 | 1.76 | 1.50 | 1.14 |
| | | % Variance | 27.10 | 15.57 | 11.00 | 9.37 | 7.13 |
| | | Cumulative % Variance | 27.10 | 42.67 | 53.67 | 63.05 | 70.17 |

deciding which sounded more acceptable and rating the verb as *manner* > *end-state* or *end-state* > *manner* as appropriate. Although this possibility cannot be ruled out entirely, it would seem extremely unlikely that the participants – undergraduate Psychology students with no training in linguistics – had explicit knowledge of the relevant semantics–syntax correlations. Participants completed the checklist in their own time, with the order of the verbs and the two statements fully randomized across participants.

2.2.3. Narrow-range semantic verb classes

In a similar way as for the broad-range rule, a new group of 15 adult participants were asked to rate, on a 10-point scale, the extent to which each verb exhibits the semantic properties listed by Pinker (1989, pp. 126–127) as diagnostic of a particular narrow-range semantic class (see Table 1). The full list of semantic properties ($N = 16$) rated by participants is given in Table 2. In order to facilitate understanding of the task, the properties were slightly simplified from Pinker's original descriptions, and cast in terms of A and B, defined for participants as follows:

A is a liquid, mass, substance, small item or a group of small items (e.g., water, paint, seeds, juice, nails, paper, clothes, bandages, stickers, bullets).

B is a container, surface or location (e.g., cup, bin, suitcase, shelf, box, sink, field, table, floor, sponge).

In principle, these ratings could be used to directly predict participants' grammaticality judgments. In practice, however, the ratio of predictors ($N = 16$) to cases would be rather high for the adults (who rated 142 verbs) and unacceptably so for the children (who rated 60 verbs). To address this problem, the 16 predictors were condensed into four via Principal Components Analysis (with no rotation), using the Eigenvalue > 1 criterion (Kaiser, 1960). The loadings of each of the original factors onto each of the five extracted narrow-range semantic predictors used in the final analysis are shown in Table 2. Note that the descriptive labels are designed merely to capture the semantic flavor of each predictor and are necessarily imprecise.

The results of the principal components analysis allow for a preliminary investigation of the semantic verb class hypothesis. If we apply a strict, literal interpretation to Pinker's (1989) proposed classes (which "line up" according to the extent with which they are compatible with the broad-range rules), we would expect to see approximately 15 small clusters, each with a large loading on just one or two of the original semantic features rated. In fact, the analysis yielded only five clusters, most with large loadings on several of the original features. However, such an interpretation would seem to be too literal; Pinker (1989) argues that the semantic class assignments proposed are not intended to be definitive, but constitute an illustration of the types of features that are relevant for classification. Thus, the finding that these semantic features form a set of relatively coherent clusters constitutes support for the spirit – if not the letter – of Pinker's (1989) proposal.

An alternative to using these semantic ratings would have been to include each verb's semantic class, as set

out by Pinker, as a categorical predictor variable. However, as noted by an anonymous reviewer, this would introduce a degree of circularity. Pinker (1989) identified classes by considering not only each verb's semantic properties but also whether each was grammatical in one or both locative constructions. Thus, provided that Pinker's intuitions regarding grammaticality correspond broadly with those of our participants, narrow-range classes could in principle "predict" participants' grammaticality judgments simply because these classes were formed on the basis of a similar set of judgments (i.e., those conducted informally by Pinker). Using (condensed) semantic ratings instead allows us to investigate whether participants' judgments are indeed sensitive to the kinds of features held to be important under Pinker's (1989) theory, whilst avoiding this potential circularity.

2.2.4. Entrenchment and pre-emption measures

Raw frequency counts were obtained from ICE-GB: a fully tagged and parsed one-million-word corpus of spoken (70%) and written (30%) British English.¹ Although larger unparsed corpora are available, it was necessary to use a fully-parsed corpus in order to allow instances of the two locative constructions to be extracted automatically. Since both hypotheses make predictions for non-alternating verbs only, we first identified verbs that are attested in only one of the two locative constructions. Verbs attested in both constructions (or neither) were assigned a score of 0 for both the pre-emption and entrenchment predictors.

The entrenchment hypothesis predicts a positive correlation between the preference for grammatical over ungrammatical verb uses and overall verb frequency, regardless of construction. Thus, the entrenchment predictor was simply the overall frequency of the verb in the corpus, $\log N + 1$ transformed. The pre-emption hypothesis predicts a positive correlation between the preference for grammatical over ungrammatical verb uses and frequency in the figure-locative construction (for figure-only verbs) and frequency in the ground-locative construction (for ground-only verbs). Thus, the pre-emption predictor was the frequency of each verb in the single construction in which it is attested² (recall that verbs attested in both were assigned a score of zero), again $\log N + 1$ transformed.

¹ Frequency counts were also obtained from the British National Corpus (both all texts and the transcribed-speech sub-corpus) and for the Manchester corpus of child-directed speech (Theakston, Lieven, Pine & Rowland et al., 2001), available on CHILDES. However, because the BNC is much larger than ICE-GB and the Manchester corpus much smaller, it was decided that the use of these measures would unfairly favour or disfavour entrenchment respectively, as compared with the pre-emption measure. Thus, the raw ICE-GB measure was used in order to allow entrenchment and pre-emption to compete on a relatively level playing field. Note, however, that this measure still favours entrenchment somewhat, as the pre-emption counts are necessarily a subset of the entrenchment counts, meaning that there is more opportunity for variation in the latter.

² The status under the pre-emption hypothesis of "short" figure- and ground-locative sentences (e.g., *Bart poured water*; *Lisa filled the cup*) is unclear. They could be seen as either (a) locative sentences with understood omitted arguments (...into the cup; ...with water), or (b) simple transitives. To be maximally generous to the pre-emption hypothesis, we decided to include short forms in our counts of each locative type.

Because the dependent measure (difference score) is a positive score for figure-only verbs and a negative score for ground-only verbs, the sign of both the entrenchment and pre-emption measure was set (after the log transformation) to positive for verbs attested in the figure-locative construction only and negative for verbs attested in the ground-locative construction only.

Because the latter are a subset of the former, the entrenchment and pre-emption counts are inevitably highly correlated ($r = 0.70, p < 0.001$). In order to allow for assessment of the respective contributions of each predictor we calculated an additional entrenchment predictor residualized against the pre-emption predictor, and vice versa.

3. Results

Mixed-effects linear regression models (see Baayen (2008) for an introduction) were fitted to the data. For

the main set of analyses, the outcome measure was a “difference score” calculated (on a verb-by-verb and participant-by-participant basis) by subtracting the rating for the ground-locative sentence from the rating from its figure-locative equivalent. In order to control for differences in the scales used by adults and children, and for any individual differences within age groups, all scores were converted into z-scores (representing the standard deviation from the mean) on a participant-by-participant basis. In addition to the fixed effects outlined below, all analyses included participant and verb as random effects. All analyses used treatment coding. Once the optimal model had been arrived at (see details below), we investigated whether adding by-participant random slopes for every fixed-effect that reached significance would yield an improved fit (for example, for the first analysis below, we added by-participant random slopes for *Manner*, *End-State* and the narrow-range semantic feature *Gluing*). In no case

Table 3

The influence of broad-range rules, narrow-range semantic verb classes, entrenchment and pre-emption on participants' preference for figure- over ground-locatives, for adults (left hand panel) and for all participants combined (right hand panel).

| | Extended set, adults only | | | | Core set, all participants | | | | | | |
|--|---------------------------|-------------|--------------|--------------|----------------------------|--------------|---------------|--------------|--------------|--------------|-----------|
| | M (β) | SE | t | P | M (β) | SE | t | P | | | |
| <i>(a) Broad-semantics model</i> | | | | | | | | | | | |
| (Intercept) | 0.04 | 0.28 | 0.15 | 0.885 | 0.39 | (0.46) | (0.83) | 0.405 | | | |
| Manner | 0.13 | 0.04 | 3.31 | 0.001 | 0.07 | 0.06 | 1.30 | 0.192 | | | |
| End State | -0.14 | 0.04 | -3.17 | 0.002 | -0.14 | 0.06 | -2.30 | 0.022 | | | |
| Verb Variance | 0.43 | | | | 0.38 | | | | | | |
| Participant Variance | 0.00 | | | | 0.00 | | | | | | |
| <i>(b) Narrow-semantics model</i> | | | | | | | | | | | |
| (Intercept) | 0.26 | 0.26 | 1.02 | 0.309 | 0.30 | 0.415 | 0.718 | 0.473 | | | |
| Manner | 0.11 | 0.04 | 2.94 | 0.003 | 0.10 | 0.055 | 1.884 | 0.060 | | | |
| End state | -0.17 | 0.04 | -4.02 | 0.000 | -0.16 | 0.055 | -2.919 | 0.004 | | | |
| S1 splattering | -0.07 | 0.05 | -1.22 | 0.223 | -0.10 | 0.065 | -1.531 | 0.126 | | | |
| S2 joining | 0.00 | 0.05 | -0.02 | 0.988 | 0.00 | 0.073 | 0.057 | 0.955 | | | |
| S3 stacking | -0.03 | 0.05 | -0.71 | 0.478 | -0.03 | 0.06 | -0.5 | 0.617 | | | |
| S4 gluing | 0.37 | 0.05 | 7.80 | 0.000 | 0.29 | 0.059 | 4.979 | 0.000 | | | |
| S5 smearing | -0.02 | 0.05 | -0.51 | 0.608 | -0.17 | 0.073 | -2.32 | 0.020 | | | |
| Verb variance | 0.29 | | | | 0.25 | | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | | | |
| <i>(c) Semantics + statistics model</i> | | | | | | | | | | | |
| (Intercept) | 0.24 | 0.22 | 1.10 | 0.273 | 0.07 | 0.34 | 0.22 | 0.827 | | | |
| Manner | 0.08 | 0.03 | 2.46 | 0.014 | 0.09 | 0.04 | 1.98 | 0.048 | | | |
| End state | -0.13 | 0.04 | -3.70 | 0.000 | -0.10 | 0.05 | -2.28 | 0.023 | | | |
| S1 splattering | -0.05 | 0.04 | -1.01 | 0.311 | -0.06 | 0.05 | -1.21 | 0.226 | | | |
| S2 joining | 0.02 | 0.04 | 0.56 | 0.576 | -0.01 | 0.06 | -0.16 | 0.873 | | | |
| S3 stacking | 0.00 | 0.04 | -0.07 | 0.943 | 0.01 | 0.05 | 0.16 | 0.877 | | | |
| S4 gluing | 0.25 | 0.04 | 5.88 | 0.000 | 0.20 | 0.05 | 3.90 | 0.000 | | | |
| S5 smearing | 0.02 | 0.04 | 0.56 | 0.577 | -0.02 | 0.07 | -0.34 | 0.733 | | | |
| Entrenchment | 0.59 | 0.08 | 7.79 | 0.000 | 0.48 | 0.09 | 5.28 | 0.000 | | | |
| Verb variance | 0.19 | | | | 0.16 | | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | | | |
| <i>(d) Statistics-only model</i> | | | | | | | | | | | |
| (Intercept) | -0.01 | 0.04 | -0.21 | 0.833 | (0.05) | (0.06) | (0.80) | 0.420 | | | |
| Entrenchment | 0.80 | 0.08 | 10.24 | 0.000 | 0.65 | 0.09 | 7.44 | 0.000 | | | |
| Verb variance | 0.52 | | | | 0.22 | | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | | | |
| Model comparisons | | | | | | | | | | | |
| (a) Broad-semantics model (df = 6) versus | | AIC | LogLik | Chi | P | df | AIC | LogLik | Chi | p | df |
| (b) Narrow-semantics model (df = 11) versus | | 9465 | -4727 | | | | 9049 | -4518 | | | |
| (c) Semantics + Statistics model (df = 12) versus | | 9419 | -4699 | 56.29 | 0.000 | 5 | 9029 | -4503 | 30.12 | 0.000 | 5 |
| (d) Statistics-only model (df = 5) | | 9368 | -4672 | 53.33 | 0.000 | 1 | 9005 | -4490 | 26.02 | 0.000 | 1 |
| | | 9402 | -4696 | 48.225 | 0.000 | 7 | 9015 | -4502 | 24.180 | 0.001 | 7 |
| Age and interactions added to Model c (df = 30) | | | | | | | 8855 | -4397 | 185.9 | 0.000 | 18 |

did this an improved fit to the data; hence we report only the simpler models with no random slopes.

3.1. Adults

The first set of analyses included all 142 verbs and ratings from adults only (as children rated only a subset of 60 verbs). The first model (termed the *Broad-semantics model*) was designed to investigate the psychological reality of Pinker's (1989) broad-range rule for the locative constructions. Ratings of the extent to which each verb denotes the (a) *manner* and (b) *end-state* of an action (corresponding to the semantic core of the figure- and ground-locative respectively) were entered as fixed effects. Both were significant predictors in the expected direction (see Table 3a). That is, the greater the extent to which the verb was rated as denoting a particular *manner*, the greater the preference for figure- over ground-locative uses (positive difference score). The greater the extent to which the verb was rated as denoting a particular *end-state*, the greater the preference for ground- over figure-locative uses (negative difference score).

The second model (termed the *Narrow-semantics model*) was designed to investigate whether adding narrow-range semantics to the first model would improve its coverage of the data. The five composite semantic features obtained from the Principal Components Analysis (see Table 2) were entered as predictors (see Table 3b). Adding these narrow-range semantic features to the model significantly improved its coverage of the data (see rows labeled *Model Comparisons*; note that smaller AIC and LogLik values indicate better model fit). Although, only one of the five features (*Gluing*) explained a significant proportion of variance by itself, it is important to note that at least some of the other semantic clusters (certainly *Stacking* and, to some extent, *Splattering*) characterize alternating verbs, whose ratings would not necessarily be expected to differ significantly from the intercept (i.e., participants' mean ratings). Nevertheless, this finding suggests that whilst the types of syntactically-relevant semantic features proposed by Pinker (1989) certainly do seem to influence verbs' argument structure privileges, the particular classes proposed (or at least the PCA clusters derived from them) do not precisely characterize how they do so.

The third model was designed to investigate whether the statistical learning (entrenchment and pre-emption) predictors explain additional variance beyond that explained by verb semantics. A complication already noted is that the entrenchment and pre-emption predictors are highly correlated ($r = 0.70$, $p < 0.001$). Thus, simply adding both to the model would not provide a reliable estimate of the unique variance explained by each. Before proceeding, we therefore investigated whether the entrenchment or pre-emption measure alone constituted a better predictor of the outcome measure. Although both entrenchment ($B = 0.80$, $SE = 0.08$, $t = 10.24$, $p < 0.001$) and pre-emption ($B = 1.05$, $SE = 0.11$, $t = 9.42$, $p < 0.001$) were excellent predictors of the difference scores, the entrenchment model (AIC = 9402, LogLik = -4696) was significantly better than the pre-emption model (AIC = 9412, LogLik = -4701) by likelihood ratio test ($\chi^2 = 9.71$, $df = 0$, $p < 0.001$).

Furthermore, whilst adding the residualized entrenchment predictor to the pre-emption model significantly improved its coverage (AIC = 9404, LogLik = -4696, $\chi^2 = 9.75$, $df = 1$, $p = 0.002$), adding the residualized pre-emption predictor to the entrenchment model did not (AIC = 9404, LogLik = -4696, $\chi^2 = 0.03$, $df = 1$, $p = 0.85$, n.s.). Thus, whilst entrenchment and pre-emption are both excellent predictors independently, when pitted against one another, only one – entrenchment – explains variance that the other cannot.

Thus, for the third model (termed the *Semantics + Statistics model*) only the entrenchment predictor was added to the *Narrow Semantics model* outlined above. This model is shown in Table 3c. Although the broad- and narrow-range semantic predictors remain significant, the entrenchment predictor explains a significant proportion of additional variance (see rows labeled *Model Comparisons*). That is, frequency in attested constructions has an effect on participants' judgments above and beyond the effect of verb semantics.³

In the introduction, we discussed the possibility that, whilst the motivation underlying particular argument restrictions may ultimately be semantic, learners can acquire these restrictions using statistical learning alone, via entrenchment/pre-emption (e.g., Stefanowitsch, 2008). If this is the case, then the statistical-learning predictors should be an almost-perfect proxy for semantics. That is, when all semantic factors are removed, a *statistics-only* model (see Table 2d) should give equally good coverage of the data. In fact, removing the semantic predictors had a significantly detrimental effect on the model's ability to explain the judgment data (see the final two *Model Comparisons*).

In summary, the best statistical model for this dataset (adult ratings of all 142 locative verbs listed by Pinker, 1989) includes both semantic measures (at the broad and narrow level) and the statistical-learning measure of entrenchment, but not pre-emption.

3.2. Developmental effects

For the full adult dataset, all semantic and statistical predictors are required to explain the pattern of grammaticality-judgment data observed. This raises the question of development: are all factors operational throughout learning and, if so, does their relative influence change over time?

With regard to the semantic verb class hypothesis, two different predictions are feasible, depending on whether or not young children are assumed to have acquired the broad-range rules. If so, one would expect to see little or no developmental progression with regard to the influence of the broad-range semantic predictors, with the influence

³ A reviewer raised the possibility that apparent frequency effects are in fact semantic effects "in disguise", and arise simply from the fact that more frequent verbs have better-learned semantics. Indeed, this is a position that we have taken in a previous paper (Ambridge et al., 2009). However, the present finding that verb frequency (entrenchment) explains additional variance beyond that explained by semantics counts strongly against this possibility.

of the narrow-range predictors increasing with age, as classes are formed (Pinker, 1989, suggests puberty as the point where a learner's narrow classes become fixed). However, another possibility consistent with the semantic verb class hypothesis is that neither the broad- nor narrow-range rules are fully acquired until some point after age 9–10, and that developmental effects will be observed at both levels.

Two different possibilities are also feasible with regard to the entrenchment/pre-emption hypotheses. On the one hand, the between-verb differences predicted under such accounts (errors are less acceptable for verbs that are of high frequency overall, or in a competing construction) may be expected to be of a similar magnitude for each age group as the ratio of high-to-low frequency tokens presumably remains constant. For example, *pour* is approximately twice as frequent as *dribble* in adult corpora, and the same is likely true for children. On the other hand, if pre-emption/entrenchment requires some threshold of occurrences, or is more sensitive to absolute than relative frequency, then the magnitude of the observed frequency effect may increase with age.

Before investigating age effects directly, we first investigated whether the effects observed above for adults, looking at all 142 verbs, held when looking at just the 60 verbs rated by all participants, including all age groups (see Table 3, right hand portion). The results were essentially unchanged, in that the *Semantics + Statistics* model yielded the best coverage of the data, with the two broad-range predictors, the same single narrow-range predictor (*Gluing*) and the entrenchment predictor significant (although the details are not shown, the optimal statistics-only model again included entrenchment but not pre-emption).

A likelihood-ratio test (see Table 3, final row) confirmed that adding age and its interactions to this model yielded significantly improved coverage of the data ($\chi^2 = 185.9$, $df = 18$, $p < 0.001$). The main set of analyses was therefore conducted for each age group independently (see Table 4).

Inspection of the semantic predictors suggests that, consistent with the second possibility outlined above, the influence of both the broad and narrow-range semantic predictors increases with age and roughly in parallel. In the optimal model (Table 4c), the Beta values for the broad-range factor of *End State* increase gradually from age 5–6 ($B = -0.07$, $SE = 0.04$) through 9–10 ($B = -0.09$, $SE = 0.05$) to adults ($B = -0.13$, $SE = 0.06$), with the younger group differing significantly from adults (Age 5–6 versus Adults \times End State: $B = 0.06$, $SE = 0.03$, $t = 2.35$, $p = 0.02$), and the older group only marginally so (Age 9–10 versus Adults \times Endstate: $B = 0.05$, $SE = 0.02$, $t = 1.77$, $p = 0.09$).⁴ For the other broad-range feature, *Manner*, the older children ($B = 0.11$, $SE = 0.04$) and adults ($B = 0.10$, $SE = 0.06$) show very similar performance (Age 9–10 versus Adults \times Manner: $B = 0.001$, $SE = 0.02$, $t = 0.43$, $p = 0.67$, n.s.), with the younger children ($B = 0.03$, $SE = 0.04$)

significantly less influenced by this feature (Age 5–6 versus Adults \times Manner: $B = -0.07$, $SE = 0.03$, $t = -2.71$, $p = 0.007$).

A very similar pattern is observed for the only significant narrow-range feature (*Gluing*), which increases in importance gradually from age 5–6 ($B = 0.12$, $SE = 0.04$) through 9–10 ($B = 0.17$, $SE = 0.05$) to adults ($B = 0.26$, $SE = 0.07$), with both child groups significantly different from adults (Age 5–6 versus Adults \times Gluing: $B = -0.14$, $SE = 0.03$, $t = -4.74$, $p < 0.001$; Age 9–10 \times Adults \times Gluing: $B = -0.09$, $SE = 0.03$, $t = -3.17$, $p = 0.002$). In fact, it is notable that this narrow-range semantic predictor explains a significant proportion of unique variance for all age groups, which is not always the case for the broad-range predictors. Furthermore, as a group, the narrow-range semantic predictors always explain additional variance above and beyond that explained by the broad-range predictors, for each age group individually.

In summary, broad- and narrow-range semantic predictors appear to show a very similar developmental trajectory. This provides support for a version of the semantic verb class hypothesis in which both broad- and narrow-range semantic properties of verbs are still being learned at age 9–10 (and possibly until puberty). A stronger version of the semantic verb class hypothesis under which children first acquire adult-like broad-range rules and only later form narrow-range semantic classes is not supported.

Adding the frequency predictor, entrenchment, also yielded significantly improved coverage for each age group (see *Model comparisons*, Model b versus Model c). Entrenchment was added as the sole frequency predictor as, for each age group, the optimal statistics-only model again included the entrenchment, but not the pre-emption predictor. As for the semantic predictors, the relative influence of entrenchment increased between age 5–6 ($B = 0.24$, $SE = 0.08$), and age 9–10 ($B = 0.53$, $SE = 0.09$) and adults ($B = 0.59$, $SE = 0.012$), with only the younger group significantly different from adults (Age 5–6 versus Adults \times Entrenchment: $B = -0.39$, $SE = 0.05$, $t = -6.44$, $p < 0.001$; Age 9–10 versus Adults \times Entrenchment: $B = -0.06$, $SE = 0.05$, $t = -1.09$, $p = 0.28$, n.s.). Thus, whilst entrenchment has a significant influence on participants' judgments at every age, this influence appears to increase between ages 5–6 and 9–10 and then to plateau.

As outlined above, both the semantic and entrenchment predictors explained significant independent proportions of variance at each age group. Thus, as one would expect, likelihood ratio tests revealed that for each age group separately (as for all ages combined), the model that offered the best coverage of the data included both the semantic- and statistical-learning predictors (see Table 4, *Model comparisons*).⁵

⁴ These comparisons are taken from the *Semantics + Statistics* Model with age and its interactions added, the parameters of which are given in the last row of Table 3. For reasons of space, this model (which contains 26 predictors) is not shown in full.

⁵ For the youngest children, the *semantics + statistics* model significantly outperformed the *statistics-only* model by log likelihood test (-1286 versus -1293), but not according to the AIC, which was slightly lower for the latter (2596 versus 2595). This is because the AIC penalizes models that contain a high number of predictors ($df = 12$ for the *semantics + statistics* model, $df = 5$ for the *statistics-only* model). Thus, whilst the *statistics-only* model is more economical, the fact remains that adding the semantic predictors nevertheless yields improved coverage of the data.

Table 4

Separate analysis for each age group.

| | Core set, adults | | | | Core set, Age 9 | | | | Core set, Age 5 | | | | | |
|---|----------------------|-------------|--------------|--------------|----------------------|-------------|--------------|--------------|----------------------|--------------|--------------|--------------|--------------|--------------|
| | <i>M</i> (β) | <i>SE</i> | <i>t</i> | <i>P</i> | <i>M</i> (β) | <i>SE</i> | <i>t</i> | <i>P</i> | <i>M</i> (β) | <i>SE</i> | <i>t</i> | <i>P</i> | | |
| <i>(a) Broad-semantics model</i> | | | | | | | | | | | | | | |
| (Intercept) | (0.56) | (0.59) | (0.94) | 0.347 | (0.17) | (0.46) | (0.37) | 0.710 | (0.28) | (0.31) | (0.92) | 0.360 | | |
| Manner | 0.09 | 0.07 | 1.22 | 0.222 | 0.09 | 0.06 | 1.69 | 0.092 | 0.02 | 0.04 | 0.65 | 0.514 | | |
| End state | -0.17 | 0.08 | -2.24 | 0.026 | -0.13 | 0.06 | -2.19 | 0.029 | -0.08 | 0.04 | -2.03 | 0.042 | | |
| Verb variance | 0.63 | | | | 0.36 | | | | 0.13 | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | 0.00 | | | | | |
| <i>(b) Narrow-semantics model</i> | | | | | | | | | | | | | | |
| (Intercept) | 0.47 | 0.53 | 0.88 | 0.382 | 0.03 | 0.43 | 0.08 | 0.938 | 0.25 | 0.31 | 0.84 | 0.404 | | |
| Manner | 0.12 | 0.07 | 1.74 | 0.083 | 0.13 | 0.06 | 2.33 | 0.020 | 0.04 | 0.04 | 0.99 | 0.324 | | |
| End state | -0.20 | 0.07 | -2.86 | 0.004 | -0.15 | 0.06 | -2.59 | 0.010 | -0.10 | 0.04 | -2.42 | 0.016 | | |
| S1 splatting | -0.12 | 0.08 | -1.39 | 0.166 | -0.11 | 0.07 | -1.70 | 0.090 | -0.06 | 0.05 | -1.15 | 0.250 | | |
| S2 joining | 0.02 | 0.09 | 0.24 | 0.808 | -0.02 | 0.08 | -0.30 | 0.763 | 0.00 | 0.05 | -0.05 | 0.958 | | |
| S3 stacking | -0.05 | 0.08 | -0.62 | 0.535 | -0.03 | 0.06 | -0.53 | 0.596 | 0.01 | 0.04 | 0.12 | 0.901 | | |
| S4 gluing | 0.38 | 0.08 | 4.98 | 0.000 | 0.27 | 0.06 | 4.42 | 0.000 | 0.17 | 0.04 | 3.82 | 0.000 | | |
| S5 smearing | -0.20 | 0.09 | -2.15 | 0.032 | -0.18 | 0.08 | -2.36 | 0.018 | -0.10 | 0.05 | -1.90 | 0.058 | | |
| Verb variance | 0.41 | | | | 0.24 | | | | 0.09 | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | 0.00 | | | | | |
| <i>(c) Semantics + statistics model</i> | | | | | | | | | | | | | | |
| (Intercept) | 0.19 | 0.45 | 0.43 | 0.665 | -0.21 | 0.34 | -0.63 | 0.530 | 0.14 | 0.28 | 0.50 | 0.614 | | |
| Manner | 0.10 | 0.06 | 1.77 | 0.077 | 0.11 | 0.04 | 2.60 | 0.010 | 0.03 | 0.04 | 0.87 | 0.387 | | |
| End state | -0.13 | 0.06 | -2.21 | 0.027 | -0.09 | 0.05 | -1.87 | 0.061 | -0.07 | 0.04 | -1.82 | 0.070 | | |
| S1 splatting | -0.07 | 0.07 | -1.04 | 0.299 | -0.08 | 0.05 | -1.43 | 0.155 | -0.04 | 0.04 | -0.84 | 0.399 | | |
| S2 joining | 0.01 | 0.08 | 0.08 | 0.937 | -0.04 | 0.06 | -0.64 | 0.523 | -0.01 | 0.05 | -0.19 | 0.847 | | |
| S3 stacking | 0.00 | 0.06 | -0.03 | 0.976 | 0.01 | 0.05 | 0.17 | 0.864 | 0.02 | 0.04 | 0.59 | 0.557 | | |
| S4 gluing | 0.26 | 0.07 | 3.89 | 0.000 | 0.17 | 0.05 | 3.26 | 0.001 | 0.12 | 0.04 | 2.77 | 0.006 | | |
| S5 smearing | -0.02 | 0.09 | -0.26 | 0.798 | -0.02 | 0.07 | -0.25 | 0.800 | -0.03 | 0.06 | -0.52 | 0.604 | | |
| Entrenchment | 0.59 | 0.12 | 4.89 | 0.000 | 0.53 | 0.09 | 5.83 | 0.000 | 0.24 | 0.08 | 3.14 | 0.002 | | |
| Verb variance | 0.28 | | | | 0.14 | | | | 0.08 | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | 0.00 | | | | | |
| <i>(d) Statistics-only model</i> | | | | | | | | | | | | | | |
| (Intercept) | 0.12 | 0.08 | 1.51 | 0.132 | -0.02 | 0.06 | -0.36 | 0.719 | -0.01 | 0.05 | -0.24 | 0.813 | | |
| Entrenchment | 0.81 | 0.11 | 7.05 | 0.000 | 0.67 | 0.09 | 7.84 | 0.000 | 0.34 | 0.07 | 5.16 | 0.000 | | |
| Verb variance | 0.37 | | | | 0.18 | | | | 0.08 | | | | | |
| Participant variance | 0.00 | | | | 0.00 | | | | 0.00 | | | | | |
| Model comparisons | | AIC | logLik | Chi | <i>p</i> | <i>df</i> | AIC | LogLik | Chi | <i>p</i> | AIC | LogLik | Chi | <i>p</i> |
| (a) Broad-semantics model (df = 6) | | 3990 | -1989 | | | | 2302 | -1145 | | | 2614 | -1301 | | |
| versus | | | | | | | | | | | | | | |
| (b) Narrow-semantics model (df = 11) versus | | 3970 | -1974 | 29.87 | 0.000 | 5 | 2286 | -1132 | 26.08 | 0.000 | 2605 | -1292 | 19.08 | 0.002 |
| (c) Semantics + Statistics model (df = 12) versus | | 3950 | -1963 | 22.89 | 0.000 | 1 | 2258 | -1117 | 30.48 | 0.000 | 2596 | -1286 | 10.47 | 0.001 |
| (d) Statistics-only model (df = 5) | | 3959 | -1974 | 23.26 | 0.002 | 7 | 2266 | -1128 | 22.31 | 0.002 | 2595 | -1293 | 12.09 | 0.098 |

3.3. Clarifying the roles of entrenchment and pre-emption

All the analyses presented up to this point have used as the dependent measure a difference score calculated – on a verb-by-verb and participant-by-participant-basis – by subtracting the rating for each ground-locative sentence (e.g., **Bart poured the cup with water*) from the rating for its equivalent figure-locative sentence (e.g., *Bart poured water into the cup*). As noted in the introduction, difference scores are commonly used in grammaticality judgment tasks of this nature, as they control for any general dispreference that participants may show for particular verbs and nouns, and any infelicities associated with particular test sentences. For example, if, for some reason, *Lisa* was felt to be a bad example of a *filling* AGENT, this would be reflected in both the figure- and ground-locative sentences for this verb.

However, the use of difference scores suffers from an important drawback when evaluating the entrenchment

and pre-emption hypotheses. Both accounts assume that repeated uses of a verb in permissible constructions (e.g., *Bart filled the bottle with juice*) contribute to an ever-strengthening inference that the use of this verb in non-attested constructions (e.g., **Lisa filled water into the cup*) is not permitted. The problem with using difference scores to test this prediction is that any apparent entrenchment or pre-emption effect could in principle be a consequence of attested uses boosting the acceptability of grammatical sentences (e.g., *Lisa filled the cup with water*) as opposed to reducing the acceptability of the ungrammatical member of the pair (e.g., **Lisa filled water into the cup*)⁶.

The final set of analyses therefore tested the entrenchment and pre-emption hypotheses, looking at ratings for ungrammatical sentences only. Focussing on the core set of verbs rated by all participants, this comprises 20

⁶ We thank an anonymous reviewer for raising this point.

figure-locative uses of ground-only verbs (e.g., **Lisa filled water into the cup*) and 20 ground-locative uses of figure-only verbs (e.g., **Lisa poured the cup with water*). The entrenchment hypothesis predicts a negative correlation between the acceptability of a particular error and the corpus frequency of the relevant verb, regardless of context. The pre-emption hypothesis predicts a negative correlation between the acceptability of a particular error and the corpus frequency of the relevant verb in the closest competing construction (i.e., figure-locative for ground-only verbs and ground-locative for figure-only verbs). Neither hypothesis makes any prediction with regard to alternating verbs. Because this analysis was conducted on raw scores, as opposed to difference scores, raw log frequency counts were used, as opposed to the counts assigned a +/- sign on the basis of their figure-only/ground-only status, used for the main analysis.

For each age group, we compared a model that included both predictors against (a) an entrenchment-only model and (b) a pre-emption-only model. The model with both predictors contained the raw entrenchment predictor and the residualized pre-emption predictor (for the purposes of comparison with the single-predictor models, such a model is formally identical to a model with a raw pre-emption predictor and a residualized entrenchment predictor). The results of this analysis are shown in Table 5.

For the adults, adding entrenchment confers a marginal improvement upon a pre-emption-only model ($p = 0.08$), but adding pre-emption confers no improvement on an entrenchment-only model ($p = 0.45$). The situation is identical for the 9–10 year olds ($p = 0.06$ and $p = 0.56$ respectively). Thus, for the two older groups, an entrenchment-only model is optimal (replicating the findings from the main analysis). For the younger children the model

containing both predictors performs significantly better than either the entrenchment-only ($p = 0.005$) or the pre-emption-only model ($p = 0.002$). However, inspection of the *Entrenchment + Pre-emption* model reveals that the residualized pre-emption predictor is working in the opposite direction to that predicted: A figure-only verb that is more frequent in the figure-locative than one would predict given its overall frequency (the residualized pre-emption measure) is *more* acceptable in the ungrammatical ground-locative construction than one would expect, given its overall frequency (and vice versa for ground-only verbs). This most likely reflects a statistical quirk arising from the residualization process. In the *Pre-emption only* model for this age group, pre-emption does not have a negative effect, but an effect of precisely zero ($M = 0.00$, $SE = 0.11$, $p = 0.98$, n.s.). Supporting this interpretation, note that for the older children and adults, the pre-emption predictor is in the predicted negative direction in the *Pre-emption-only* model, but the residualized version flips (though is not significant) in the *Entrenchment + Pre-emption* model.

In summary, the entrenchment hypothesis was supported for all age groups: The greater the overall frequency of a verb, the less acceptable it is rated in non-permitted constructions. The effect of pre-emption – for the older two groups – requires a more nuanced explanation. To be sure, the greater the frequency of a ground-only verb in the ground-locative construction, the less acceptable it is rated in the figure-locative construction (and vice versa), in accordance with the predictions of the pre-emption hypothesis. However, to the extent that entrenchment and pre-emption can be dissociated statistically, there is no evidence that occurrence in one of the two locative constructions (pre-emption) has any effect over and above occurrence in general (entrenchment).

Table 5
Clarifying the roles of entrenchment and pre-emption.

| | Core set, adults | | | | Core set, Age 9 | | | | Core set, Age 5 | | | | | |
|---|------------------|---------------|----------------|--------------|-----------------|-------------|---------------|---------------|-----------------|--------------|---------------|---------------|------|-------|
| | $M(\beta)$ | SE | t | P | $M(\beta)$ | SE | t | P | $M(\beta)$ | SE | t | P | | |
| <i>Entrenchment + Pre-emption</i> | | | | | | | | | | | | | | |
| (Intercept) | -0.34 | 0.13 | -2.53 | 0.011 | -0.42 | 0.10 | -4.26 | 0.000 | -0.25 | 0.08 | -3.26 | 0.001 | | |
| Entrenchment | -0.35 | 0.15 | -2.32 | 0.021 | -0.32 | 0.12 | -2.78 | 0.006 | -0.19 | 0.09 | -2.17 | 0.030 | | |
| Pre-emption (residualized) | 0.28 | 0.36 | 0.77 | 0.442 | 0.17 | 0.28 | 0.61 | 0.546 | 0.60 | 0.21 | 2.83 | 0.005 | | |
| Verb variance | 0.27 | | | | 0.00 | | | | 0.00 | | | | | |
| Participant variance | 0.04 | | | | 0.12 | | | | 0.04 | | | | | |
| <i>Entrenchment only</i> | | | | | | | | | | | | | | |
| (Intercept) | -0.36 | 0.13 | -2.75 | 0.006 | -0.44 | 0.10 | -4.51 | 0.000 | -0.29 | 0.08 | -3.60 | 0.000 | | |
| Entrenchment | -0.32 | 0.15 | -2.21 | 0.027 | -0.31 | 0.11 | -2.74 | 0.006 | -0.13 | 0.09 | -1.42 | 0.156 | | |
| Verb variance | 0.27 | | | | 0.00 | | | | 0.00 | | | | | |
| Participant variance | 0.04 | | | | 0.12 | | | | 0.05 | | | | | |
| <i>Pre-emption only</i> | | | | | | | | | | | | | | |
| (Intercept) | -0.47 | 0.11 | -4.10 | 0.000 | -0.53 | 0.08 | -6.33 | 0.000 | -0.37 | 0.07 | -5.43 | 0.000 | | |
| Pre-emption | -0.27 | 0.18 | -1.50 | 0.134 | -0.28 | 0.14 | -2.01 | 0.045 | 0.00 | 0.11 | 0.03 | 0.975 | | |
| Verb variance | 0.29 | | | | 0 | | | | | | | | | |
| Participant variance | 0.04 | | | | 0.13 | | | | | | | | | |
| Model comparisons | | | | | | | | | | | | | | |
| Entrenchment + Pre-emption (df = 6) | | AIC | LogLik | Chi | P | df | AIC | LogLik | Chi | P | AIC | LogLik | Chi | P |
| versus | | 2856.4 | -1422.2 | | | | 1752.6 | -870.3 | | | 1863.8 | -925.9 | | |
| (a) Entrenchment only (df = 5) | | 2854.9 | -1422.5 | 0.57 | 0.450 | 1 | 1750.9 | -870.5 | 0.33 | 0.565 | 1869.6 | -929.8 | 7.79 | 0.005 |
| (b) Pre-emption only (df = 5) | | 2857.5 | -1423.7 | 3.09 | 0.079 | 1 | 1754.1 | -872.0 | 3.49 | 0.062 | 1871.7 | -930.8 | 9.86 | 0.002 |
| Entrenchment-only versus Pre-emption only | | | | 0.00 | 1.000 | 0 | | | 0.00 | 1.000 | | | 0.00 | 1.000 |

3.4. Summary

Although these findings are somewhat complex in their fine-grained detail, a simple pattern emerges. For each age group individually, for all participants combined, and whether looking at the core or extended verb set, optimal coverage of participants' grammaticality judgment data is achieved by a model that includes broad-range semantic rules, narrow-range semantic verb classes and entrenchment, with no additional role for pre-emption. In other words the statistical-learning predictor of entrenchment always adds predictive validity to a model containing only semantic predictors, and vice versa. Developmental effects were also observed, whereby the influence of all three predictors increases with age.

4. Discussion

The present study investigated how children learn that some verbs may appear in the figure-locative but not the ground-locative construction (e.g., *Lisa poured water into the cup*; **Lisa poured the cup with water*), with some showing the opposite pattern (e.g., **Bart filled water into the cup*; *Bart filled the cup with water*), and others appearing in both (*Lisa sprayed water onto the flowers*; *Lisa sprayed the flowers with water*). Grammatical acceptability judgments were obtained for the use of each of 142 locative verbs (60 for children) in each sentence type. For each age group separately, and for all age groups combined, the judgment data were best explained by a model that included ratings of the extent to which each verb exhibits both the broad-range and narrow-range semantic properties of the figure- and ground-locative constructions (Pinker, 1989) and the statistical-learning measure of overall verb frequency (entrenchment; Braine & Brooks, 1995), though not frequency in each of the two locative constructions (pre-emption; Goldberg, 1995). Importantly, removing either the semantic or entrenchment predictors had a significantly detrimental effect on the model's ability to account for the pattern of judgment data for every age group.

This suggests the need for a learning mechanism that incorporates verb semantics – at both the broad- and narrow-range levels – and statistical learning. In this section we outline a possible new integrative account of this nature, and contrast it with existing proposals (in particular, that of Pinker (1989)).

4.1. Links between semantics and syntax

We begin by setting out the evidence for links between semantics and syntax (a) in general and (b) with respect to the locative constructions, before discussing two classes of proposals for how these links can be accounted for theoretically (*lexicalist* and *construction-based* approaches).

A number of recent studies have provided evidence not only that semantics–syntax links exist, but that they are exploited by learners. Such findings provide evidence for Gleitman's (1990) *syntactic bootstrapping* hypothesis, under which children use the syntactic frames in which verbs occur to learn something about their meaning. For

example, Naigles (1990) used a preferential-looking paradigm to show that children aged as young as 2;1 were able to infer that a novel verb presented in a transitive sentence (e.g., *The duck is gorpung the bunny*) refers to a causal action (e.g., the duck pushing the bunny into a squatting position), as opposed to a synchronous action (e.g., the duck and the bunny making arm circles). Similar findings were reported by Naigles and Kako (1993), Fisher (1996), Naigles (1996), Hirsh-Pasek, Golinkoff and Naigles (1996), Bavin and Growcott (2000) and Kidd, Bavin, and Rhodes (2001), though see Pinker (1994) for discussion and some important caveats.

Indeed, the cues to meaning supplied by verb syntax are so powerful that, in cases where the meaning of a verb is less than fully compatible with the meaning of the construction, children are sometimes willing to bend the meaning of even a well-known verb to be consistent with that of the construction. For example, when asked to enact a sentence such as **The zebra goes the lion*, children younger than 9;0 typically modify the meaning of *go* (to mean something like *take* or *accompany*) rather than violating the two-participant causative semantics of the transitive construction by having the zebra *go* (Naigles, Fowler, & Helm, 1992; Naigles, Gleitman, & Gleitman, 1993).

It has long been acknowledged in the literature that the figure- and ground-locative constructions exhibit subtly different semantic properties (e.g., Anderson, 1971), though there have been many different attempts to delineate precisely what these properties are. Levin and Rappaport Hovav (1991, p. 146) are essentially in agreement with Pinker (1989) that the figure- and ground-locative variants denote (a) "simple change of location" with "the manner/means lexicalized" and (b) "change of state brought about by means of the change of location" respectively.

Beavers (2010), however, argues that the *manner-* (figure-locative) versus *end-state* (ground-locative) distinction is not quite the right one. The *cut/slice* alternation is similar in many respects to the *figure/ground*-locative alternation, except that in some cases it is the figure, not the ground, that undergoes a greater degree of state change (e.g., *John scratched the diamond against the glass*). Beavers (2010, p. 832) argues that the facts of both alternations can be accommodated by assuming that construction choice is determined by whether the figure- or the ground- is more *holistically* affected "relative to the alternating argument".

In contrast, Jeffries and Willis (1984, p. 721) argue that whilst the holism constraint on the direct object accounts for the behavior of *cover*-type verbs (e.g., *John covered the slice of bread with jelly*/**John covered jelly on the slice of bread*), it falls down for verbs like *clean*, *clear*, *drain* and *empty* (which are clearly locative verbs at least to some degree). Jeffries and Willis (1984, p. 721) point out that both the *ground-* and *figure-* object versions of sentences such as *He drained the pond of water*/*He drained water out of the pond* imply that *all* of the water (figure) has been removed from *all* of the pond (ground). These authors do not argue that the notion of holism is irrelevant to locative choice, but that it is simply one of a number of semantic factors that must be taken into account. This would seem to be a reasonable position. Most likely, the figure- and ground-locative

constructions differ along multiple semantic dimensions simultaneously, with each of the approaches outlined above capturing just one.

The question that now arises is how best to capture these semantics–syntax links (both for the locative constructions and more generally) within a linguistic framework. There are essentially two possibilities. Under lexicalist proposals (e.g., Pinker, 1989; Levin & Rappaport Hovav, 1991), each individual verb has a “basic sense” that denotes either a particular manner or a particular end-state, and that hence yields a figure- or ground-locative respectively. Non-alternating verbs have only this basic sense. An alternating verb is either content (figure) or container (ground) oriented in its basic sense (see Tables 1 and 2) but can take on an “extended sense” by means of a “gestalt shift”. For example, *spray* in its basic sense denotes an event in which a figure (e.g., *paint*) is caused to move in a particular manner (yielding a figure-locative), but can take on an extended sense whereby the ground (e.g., *the wall*) is caused to undergo a state-change (yielding a ground-locative). Note that, as discussed in the introduction, this transformation does not occur at the level of syntax. Rather, one lexico-semantic structure is transformed into the other, with innately specified semantics–syntax linking rules spelling out the argument structure for the verb.⁷ The most important linking rule for our present purposes is that which spells out the [THEME] as the [DIRECT OBJECT]. The alternation rule specifies whether the figure or ground is construed as the theme, and hence is expressed as the direct object (e.g., *Bart sprayed paint onto the wall*/*Bart sprayed the wall with paint*).

Under construction-based approaches (e.g., Goldberg, 1995; Iwata, 2008) it is not only individual verbs but also the figure- and ground-locative constructions themselves that are associated with *motion in a particular manner* and *change of state* respectively.⁸ When a verb is used in a particular construction, both the verb and the construction contribute to the overall meaning (a process known variously as “fusion” [Goldberg, 1995, p. 50], “unification” [Kay and Fillmore, 1999] or “elaboration” [Langacker, 2000]). Thus, the subtle difference in meaning between figure-/ground-locative sentences with the same verb (e.g., *Bart sprayed paint onto the wall*/*Bart sprayed the wall with paint*) is attributed to the different meanings associated with

the two constructions. Construction-based accounts do not posit alternations as such; “alternating” verbs do not have two different senses, but a single central sense that is compatible with the meaning of both constructions.

Lexicalist and construction based approaches have a number of differences (see also Goldberg, 1995), but also share some similarities. Importantly, both assume the existence of some kind of syntax–semantics links in general, and between (a) the figure-locative and manner and (b) the ground-locative and end-state in particular, regardless of how this is precisely characterized and implemented (i.e., broad-range rules versus construction semantics). Both also share the assumption that the extent to which a particular verb may be used grammatically in a particular construction is related to the degree of semantic compatibility between the two. Again, however, the precise implementation of this assumption is different; framed in terms of compatibility between narrow-range classes and broad-range rules under lexicalist accounts, and between individual verbs and construction semantics under construction-based accounts.

A matter of some debate is whether or not lexicalist accounts can explain the graded, probabilistic nature of participants’ judgments. As noted in the introduction, one possible interpretation of Pinker’s account (1989) is that the assumption that narrow classes differ as to their compatibility with the broad-range rule leads to the prediction of graded judgments (though arguably only between members of different narrow-range classes). Furthermore, these narrow-range classes are formed on the basis of an input-based generalization procedure which itself could plausibly be seen as probabilistic in nature. A stronger interpretation, however, is that the outcome of this (perhaps probabilistic) generalization procedure is a deterministic system whereby the use of a particular verb in a particular construction is either grammatical or ungrammatical, with no room for degrees of (un)grammaticality.

Whichever interpretation is closer to that originally envisaged by Pinker (1989),⁹ it is clear that an explanation of the present findings requires an account under which the notion of semantic compatibility is indeed probabilistic rather than all-or-nothing in nature, and this is reflected in the learning mechanism. One possible account of this nature is outlined shortly.

4.2. Statistical learning

Although the present article has somewhat emphasized the role of semantics, it is important to bear in mind that, in every analysis, the entrenchment measure explained additional variance when added to a model containing just the semantic predictors. Indeed, this should not be a

⁷ These linking rules are also an important part of Pinker’s (1984) *semantic bootstrapping* hypothesis; a proposal for how children acquire the syntax of their language. For example, a child who hears *The man rolled the ball* and observes that the rolling action has an AGENT and a PATIENT can use the innate rules linking AGENT → SUBJECT and PATIENT → OBJECT to infer that English uses [SUBJECT] [VERB] [OBJECT] word order.

⁸ In fact, Goldberg (1995) argues that the figure- and ground-locative constructions are instantiations of the higher level *caused-motion construction* and *causative-construction plus ‘with’ adjunct* respectively. This captures similarities between figure-locative sentences (e.g., *John loaded hay onto the wagon*) and examples such as *Frank kicked the dog into the bathroom* or *Sam ordered him out of the house* and between ground-locative sentences and simple transitives (e.g., *John loaded the wagon*). Other constructional approaches (e.g., Iwata, 2008) argue that these high-level constructions are too general to account for the distributional facts. However, this debate is not important for our present purposes, as all construction-based approaches share the general assumption of a link between *manner* and figure-ground order and *end-state* and ground-figure order.

⁹ Whilst the extent to which Pinker’s (1989) account can accommodate degrees of (un)grammaticality is debatable, Pinker himself certainly acknowledges the existence of borderline ungrammatical utterances (or, “Haigspeak”). Many of these seem to be cases where the “error” is made deliberately for a comic or other special effect (“*We’re gonna splash and we’re gonna spin ya. *We’re gonna scream and *we’re gonna grin ya*”; Pinker, 1989, p. 295). Iwata (2008, p. 84) also notes the following corpus example: “Fill water into the tank until the indicator reads that it is $\frac{3}{4}$ to full and close the cap”.

surprise, as many studies have found an effect of entrenchment when holding semantics constant (Ambridge et al., 2008; Brooks et al., 1999; Stefanowitsch, 2008; Theakston, 2004; Ambridge et al., 2009; Ambridge et al., 2011).

What of pre-emption? Why did the present study fail to find such an effect, when it has been observed in numerous previous studies (e.g., Boyd & Goldberg, 2011; Brooks & Zizak, 2002; Brooks et al., 1999; Goldberg, 2011). In fact, there is no discrepancy; the present study *did* find a pre-emption effect, at least for older children and adults. For these two groups, a significant negative correlation was observed between frequency in the figure-locative construction and the acceptability of ungrammatical ground-locative utterances (and vice versa), exactly as predicted by a pre-emption account. However, whilst pre-emption did have a significant negative effect on the acceptability of errors, it did not have a significant additional effect over and above entrenchment. Ungrammatical ground uses of a particular verb (e.g., **Lisa poured the cup with water*) were, as predicted blocked by figure-locative uses of that verb (e.g., *Bart poured water into the cup*) – and vice versa – but no more so than by any use of that verb (e.g., *The rain poured down*). The studies cited above as evidence for pre-emption did indeed provide evidence for this mechanism, but not for the claim that it has an effect over and above entrenchment.

4.3. Semantics versus statistics

Particularly important in teasing apart the effects of semantics and statistics are the statistical learning studies of Wonnacott et al. (2008) and Wonnacott (2011) (though these studies did not allow for dissociation of entrenchment and pre-emption). Wonnacott and colleagues found that adults and children demonstrated an entrenchment/pre-emption effect when acquiring novel constructions that had no semantics. Thus, it is clear that an account where apparent frequency effects are actually semantic effects in disguise, arising simply from the fact that more-frequent verbs have better-learned semantics (e.g., the account proposed by Ambridge et al. (2009)) is not tenable. This conclusion is corroborated by the findings of the present study, where an entrenchment effect was observed even after factoring out verb semantics.

On the other hand, it is equally clear that the converse account, whereby apparent semantic effects are actually frequency effects in disguise, also fails to account for the data. Both the present study and previous work (e.g., Ambridge et al., 2008, 2009; Ambridge et al., 2011, in press) have demonstrated semantic effects over and above frequency effects. Thus, it cannot be the case that, as argued by Stefanowitsch (2008), restrictions that are ultimately semantically motivated are nevertheless acquired by learners on a purely statistical basis.

What is needed, then, is an account that yields both semantic and statistical-learning effects. One possibility is simply to adopt Pinker's (1989) approach, perhaps with some additional modification. As discussed above, since Pinker's (1989) class-formation process operates on the basis of exposure to the input, it would be plausible to assume that lower frequency verbs take longer to be

assigned to the appropriate semantic class, and hence are more susceptible to error (though Pinker does not discuss this possibility explicitly). Whilst this would yield a frequency (entrenchment) effect, it would not explain why such an effect is observed over and above effects of verb semantics. One possibility would therefore be to add an additional entrenchment mechanism onto Pinker's (1989) account; indeed essentially this solution has been previously proposed (Tommasello, 2003, p. 182).

We would suggest, however, that this solution is less than fully satisfactory, as it fails to explain precisely how the different processes interact: what is the nature of the learning mechanism such that it yields effects of semantics and entrenchment? In the following section, we integrate a number of previous proposals to yield one possible account of such a mechanism. We certainly do not wish to claim that the present findings provide support for our proposal over the account of Pinker (1989), or even that they could in principle be used to decide the issue. We do, however, see the accounts as making different predictions that should be tested in subsequent research.

4.4. Integrating semantics and statistics: a new account

The account that we are proposing here is construction-based rather than lexicalist in nature (though we would not wish to claim that only this type of account is capable of accounting for the observed data). We should also note that this account is “new” only to the extent that it combines elements of various previous approaches in a novel way.

As under all construction-based accounts, children acquire constructions (including both the figure- and ground-locative constructions) by abstracting across utterances in the input that instantiate them. Each individual slot (and, most importantly, the [VERB] slot) probabilistically exhibits the semantic properties shared by the items that have appeared in this position. Thus, roughly speaking, the [VERB] slot of the figure- and ground-locative constructions will be associated with the semantics of *motion in a particular manner* and *change-of-state* respectively. It is important to note that any such characterization is necessarily imprecise because each slot exhibits a complex constellation of semantic micro-features, which will differ slightly from speaker to speaker according to the verbs that have appeared in this construction in the individual's input.

We share with Bate's and Macwhinney's (1987) competition model the assumption that “in production, forms compete to express underlying intentions or functions” (MacWhinney, 2004, p. 71). In particular, we assume that different constructions (forms) compete for the right to express a particular meaning that the speaker has in mind (see also Bowerman, 1981). For example, assume that the speaker's “message” is roughly that *Lisa poured water into a cup, causing it to become full*, and that she intends to use the lexical items *Lisa*, *fill*, *the* + *cup* and *water*. Let us assume that all the constructions in the speaker's inventory that express at least some part of this message (i.e., that meet some criterion for **relevance**) compete for activation. This would include not only the ground-locative (which

would yield *Lisa filled the cup with water*) and the figure-locative (**Lisa filled water into the cup*), but also the simple transitive-causative (*Lisa filled the cup*), the intransitive particle construction (e.g., *The cup filled up*), the periphrastic causative (*Lisa made the cup fill with water*), and so on.

With one further assumption, this competition model can yield entrenchment and pre-emption effects. The assumption is that any given verb activates a number of different constructions in memory, with activation proportional to the frequency with which it has been previously encountered in each. This assumption is well supported by studies of adult sentence processing in which (for example) different verbs preferentially activate the simple transitive and complement clause constructions, leading to different readings (e.g., Fodor, 1978; Ford, Bresnan, & Kaplan, 1982; Clifton, Frazier, & Connine, 1984; Boland & Tanenhaus, 1991; Ferreira & Clifton, 1986; Gahl, 2002; Gahl, Jurafsky, & Roland, 2004; Jennings, Randall, & Tyler, 1997; Lapata, Keller, & Walde, 2001; MacDonald, 1994; Trueswell, Tanenhaus, & Kello, 1993):

The doctor remembered... the patient > that the patient was sick
 The doctor suspected... that the patient was sick > the patient

It is a small step to assume additionally that the verb that a speaker plans to use activates candidate constructions in proportion to the frequency with which it has appeared in each (i.e., that the verb has a finite amount of activation that it shares out between different constructions).

This effect of **verb-in-construction** frequency yields both pre-emption and entrenchment effects. Pre-emption effects arise because every encountered occurrence of (for example) *fill* in the ground-locative construction increases the extent to which *fill* preferentially activates the ground- versus figure-locative construction in production. Entrenchment effects arise in the same way. The more often *fill* has been encountered in any construction (e.g., ground-locative, transitive causative, intransitive with *up*, periphrastic causative), the greater the extent to which *fill* will activate all of these constructions in production, at the expense of the (ungrammatical) figure-locative. In the present study, an effect of pre-emption was observed, but it did not explain variance beyond that accounted for by entrenchment. The explanation for this finding under the present account is that the use of *fill* in any construction boosts the link between *fill* and this construction, hence increasing the probability that, when a speaker is planning an utterance with *fill*, it is this construction – and not the figure-locative – that is activated.

It would also seem likely (although it is not crucial for the present account) that constructions with higher overall frequency have a higher resting activation level. This would explain why, all other things being equal, speakers generally select the higher frequency construction of the alternatives available (e.g., the transitive causative over the periphrastic causative).

The final step is to integrate these frequency effects with semantic effects. All that is required here is the

assumption that the mechanism that selects between the competing constructions is sensitive not only to *verb-in-construction* frequency but to the semantic *fit* between individual verbs and individual construction slots.¹⁰ Thus, for example, *fill* is dispreferred in the figure-locative not only because it has occurred frequently in the ground-locative and other constructions (pre-emption/entrenchment), but also because its semantics (roughly, “cause to become full”) are less than fully compatible with those of the [VERB] slot in the figure-locative construction (roughly, “cause to move in a particular manner”). That is, the present account shares with that of Langacker (2000, p. 17) the assumption that “an expression is ill-formed to the extent that any [construction slots] involve extension rather than elaboration” (see also Bowerman’s, 1981, notion of “harmony” between verb and construction semantics). Because semantic fit is probabilistic, as opposed to all-or-nothing, in nature, it could potentially yield the fine-grained semantic effects observed in the present study.¹¹

An advantage of this mechanism is that it can potentially account for both semantic and statistical-learning effects observed for other alterations and constructions, including (a) the intransitive-inchoative/transitive-causative, (b) the prepositional-object/double-object dative and (c) the reversative *un*-prefixation construction. As outlined in the introduction, all appear to exhibit particular semantic constraints such as (a) *internal versus direct-external causation*, (b) *causing to go versus causing to have* and (c) *circular motion, hand-movements* (see Bowerman, 1981; Pinker, 1989; Whorf, 1956). It may even be possible to extend this proposal to account for overgeneralization errors based on a mismatch between phonological and pragmatic properties of particular items and particular construction slots (e.g., Ambridge et al., 2011, p. 318).

Note that an important, and perhaps controversial, aspect of the present proposal is its assumption that there exist few (perhaps even no) true exceptions to semantically-motivated argument-structure generalizations (or those based on phonology or pragmatics). Apparent counterexamples simply reflect the failure of investigators to narrow down the relevant semantic, phonological or pragmatic criteria with sufficient precision. This assumption is shared by Pinker (1989, p. 103), who makes explicit his aim to “leave no negative exceptions”. In contrast, “pure” entrenchment/pre-emption accounts generally assume that the grammar contains some arbitrary negative exceptions that must be learned on an entirely statistical basis.

¹⁰ Because it emphasises the *fit* between properties of *items* and (construction) *templates*, we suggest the “fit” account as an acronym for this proposal.

¹¹ It is important to emphasise that our goal has not been to argue against a straw-man position that “semantics is not important” when determining the grammaticality of particular argument structure generalizations, but rather to begin to build an account of precisely *how* semantics exerts its effects. Our claim is not simply that “speakers avoid utterances that make no sense (e.g., **John seemed the flowers with water*)”, though this is of course true. The very paradox that this research aims to address is that certain utterances (e.g., **John poured the flowers with water*) are deemed less than fully grammatical despite the fact that, at a broad-brush level, they *do* make sense; this is precisely why children produce them (and it would be a strange listener indeed who was unable to interpret a speaker’s request to “Pour the flowers with water”).

4.5. Development

The construction-based account outlined here and the lexicalist account of Pinker (1989) offer similar accounts of development; though as we will see shortly, they do potentially make some different predictions.

Both accounts assume that one source of children's errors is non-adult-like verb meanings. Before the correct meaning of a verb has been honed, children may fail to appreciate any incompatibility between its meaning and that of the construction (under the present account) or assign it to an inappropriate semantic class (under the account of Pinker, 1989). Evidence in support of this prediction comes from the study of (Gropen, Pinker, Hollander and Goldberg (1991a,b; see also Gentner, 1975, 1978, 1982). These authors found that children who produced overgeneralization errors with *fill* (e.g., **Lisa filled water into the cup*) judged pouring events where a glass ended up only $\frac{3}{4}$ full to be perfectly good examples of "filling". This suggests that, for these children, the meaning of *fill* was something closer to the adult meaning of *pour* and, consequently, perfectly compatible the meaning of the figure-locative construction (or with a figure-locative class).

Both accounts also share the assumption that errors will cease gradually as children refine their knowledge of verb meanings; for Pinker (1989) because this will result in correct semantic-class assignment; for the present account because the relative compatibility between the verb and different construction slots will become more readily apparent. Both also allow for the possibility that children will make errors as a result of (a) lexical gaps (e.g., using *fill* instead of *pour* in a figure-locative construction if they have yet to learn the verb *pour*) and (b) immature metalinguistic abilities. Pinker (1989, p. 295) uses the example of an advertisement for an amusement park, where an overgeneralization *We're gonna spin ya...we're gonna grin ya...* is "excused by the carefree tone", arguing that children will be "oblivious to these nuances".

The present account assumes that children will also make errors as a result of non-adult-like knowledge of construction – as opposed to verb – semantics. The fine-grained semantic properties of construction slots (most importantly the [VERB] slot) are built up slowly and gradually over time; indeed learning presumably continues into adulthood. Thus, an additional source of error is children's failure to have yet acquired adult-like knowledge of the properties of a particular construction slot (e.g., that the VERB slot in the figure-locative construction has the semantic property of *manner*, with which *fill* is incompatible).¹² Overgeneralization errors cease as children refine and strengthen their knowledge of the properties of individual construction slots. Pinker's (1989) account also shares this prediction, provided that it possible to reconceptualise learning the semantic properties of each [VERB] slot as learning which particular narrow-range classes are semantically consistent with which broad-range rule (i.e., *manner*, *end-state*, or both).

There is, however, one area where the two accounts would appear to make different predictions. Under Pinker's (1989) account, the broad-range rules are acquired early (on the basis of innate semantics–syntax linking rules), with the narrow-range rules acquired only later. Under the present account, no distinction is made between broad- and narrow-range semantics; learning of all semantic properties of both verbs and construction slots begins as soon as verbs and constructions start to be acquired.

Although we probably did not test young enough children to directly examine the prediction that children acquire the broad-range rule before the narrow-range classes, the present study provides one piece of suggestive evidence against this claim. Recall that the influence of the both broad- and narrow-range rules increased in tandem with age. Furthermore, in the optimal model for the youngest children, a narrow-range predictor, but no broad-range predictor, was associated with a significant effect.

Future studies should test younger children, perhaps using production or comprehension as opposed to judgment paradigms, to investigate whether there is any evidence for the claim that children acquire broad-range rules before narrow-range semantic properties. Such studies would help to mediate between the present account and that of Pinker (1989).

Future studies, again focussing on younger children, should test the prediction that children make overgeneralization errors as a consequence of non-adult-like knowledge of verb and construction semantics. Although both the present account and that of Pinker (1989) make this prediction, virtually no studies have attempted to test it directly. One exception, with regard to verb semantics, is the study of Gropen et al. (1991a, 1991b), which suggests a suitable paradigm for such investigations in the future. With regard to construction semantics – where we are aware of no relevant studies – an experimental approach that may be useful is the novel-construction-learning paradigm of Goldberg and colleagues (Boyd, Gottschalk, & Goldberg, 2009; Casenhiser & Goldberg, 2005; Wonnacott, Boyd, Thomson, & Goldberg, submitted for publication). These studies involve teaching English-speaking adults and children novel argument structure constructions characterized by non-English word order (and sometimes case-marking morphemes) associated with a particular meaning (e.g., appearance). A particular advantage of this paradigm is that it is possible to invent constructions that violate Pinker's (1989) proposed linking rules (see Casenhiser & Goldberg, 2005, p. 505). This allows researchers to investigate children's ability to form semantically-governed construction slots in the absence of assistance from linking rules (or familiar English constructions).

In conclusion, the new hybrid mechanism that we have begun to outline here seems to offer a plausible account of the empirical grammaticality judgment data, although much work remains to be done with regard to differentiating the proposal from that of Pinker (1989), particularly with regard to explaining children's development. The major finding of the present study is that both semantic and statistical-learning effects are psychologically real, and that both explain independent variance in participants' judgments at all ages. Thus, whether or not either of the

¹² Another possibility, whether or not the child is sensitive to sub-optimal semantic fit, is that she has yet to learn a construction with a more suitable slot (e.g., she may have learned the figure- but not the ground-locative).

accounts that we have discussed here turns out to be correct, any successful account of the formation and restriction of linguistic generalizations will have to include a role for both factors. Future studies designed to test further predictions of these accounts, and of any subsequent rival accounts that yield the observed effects, will bring the field closer to an understanding of how children avoid overgeneralization errors whilst retaining the capacity for productivity: the defining characteristic of human language.

Appendix A. Grammaticality judgment training sentences

Grammaticality-judgment training trials and “typical” scores reported to the participants (based on overall mean ratings in the study of Ambridge et al., 2008). Children completed two more training trials than the adults, to provide them with extra practice.

| | |
|--|--------|
| The frog caught the fly (Experimenter completes) | 5 |
| His teeth man the brushed (Experimenter completes) | 1 |
| The cat drank the milk (children only) | 5 |
| The dog the ball played with (children only) | 1 |
| The man tumbled Bart into a hole | 2 or 3 |
| The magician vanished Bart | 2 or 3 |
| The funny clown giggled Lisa | 1 or 2 |

Appendix B. Training examples for the semantic rating task

Participants were given the following instructions:

These examples are designed to show you how to use the scale to rate the statements appropriately.

The word describes the particular manner/way in which the action occurs (i.e., if the manner/way changes, this word is no longer an appropriate description of the event).

| Verb | Rating |
|-------|--------|
| Go | 1 |
| Kill | 1 |
| Walk | 10 |
| Run | 10 |
| Hurry | 5 |

Consider the verb *go*. This does NOT describe a particular manner/way in which the action occurs. For example, a person might go to the shops by walking, running, jogging, driving, etc. If the manner changed (e.g., from walking to running) *go(ing)* would still be an appropriate description. Therefore *go* does not describe a particular manner. Thus, the appropriate rating is 1/10 (i.e., strongly DISAGREE with

the statement that the verb describes the manner/way in which the action occurs). Similarly, *kill* does not describe a particular manner in which the action occurs (*kill* is still an appropriate description whether it is by stabbing, shooting, poisoning, etc.): Rating = 1/10.

Consider now the verb *walk*. This DOES describe a particular manner/way in which the action occurs. That is, if the manner/way changes (for example from walking to running), then the verb *walk* is no longer an appropriate description of the relevant event. Therefore *walk* does describe a particular manner. Thus, the appropriate rating is 10/10 (i.e., strongly AGREE with the statement that the verb describes the manner/way in which the action occurs). For exactly the same reasons, *run* is a manner verb, and a score of 10 is again appropriate.

Hurry is an example of a verb for which a rating on the middle of the scale (around 5/10) would be appropriate. On the one hand, *hurry* does not specify a particular manner to as great an extent as a verb such as *walk*: One can *hurry* whilst walking, jogging, or doing something else altogether (e.g., writing an essay or eating a meal). On the other hand, *hurry* does specify a manner (“quickly”) to a greater extent than a verb such as *kill* does: If the action changed from walking quickly to walking slowly (or eating quickly to eating slowly) then the verb *hurry* would no longer be an appropriate description of the relevant event.

The word describes the end-state of an action (i.e. if the relevant end-state is not arrived at, this word is no longer an appropriate description of the event).

| Verb | Rating |
|------|--------|
| Kill | 10 |
| Open | 10 |
| Walk | 1 |
| Heat | 5 |

A verb describes the end-state of an action if some person or thing ends up in a particular state as a result of the action. Consider the verb *kill*. An important part of the meaning of the verb is that, as a result of the action, a person arrives at a particular end-state (i.e., being dead). If the relevant end-state (death) is not arrived at then *kill* would no longer be an appropriate description of the relevant event (if one shoots a person repeatedly in the head but, somehow, they do not end up dead, then the verb *kill* would no longer be an appropriate description of the relevant event). Therefore *kill* certainly does describe the end-state of an action, and a rating of 10/10 (strongly agree) is appropriate. *Open* is another example of an end-state verb. If the door, jar, etc. does not arrive at the end-state of being open, the verb is not an appropriate description of the relevant event.

Walk is an example of a verb that does not describe a particular end-state. Of course, a person who walks MIGHT end up in a particular state (e.g., exhausted), but this is not part of the meaning of the verb. *Walk* is still an appropriate description of any action that involves the relevant movement of the legs, whether or not any particular end-state is arrived at. Thus, *walk* certainly does not describe

the end-state of an action, and a rating of 1/10 (strongly disagree) is appropriate.

Heat is an example of a verb for which a rating of around 5/10 would be appropriate. On the one hand, it does not imply a particular end-state to as great an extent as *kill* does, as something that is heated may not end up hot (e.g., soup that is heated may end up only slightly warm; ice that is heated may end up as cold water). On the other hand, *heat* certainly implies an end-state to a greater extent than a verb such as *walk*. Something is only heated if it ends up at least somewhat hotter than it was.

On the sheet overleaf you will see a list of 142 verbs. Each of these verbs is slightly different in meaning. Your task, for each verb, is to rate the accuracy of these two statements about the verb's meaning on a scale from 1 to 10 where 10 = strongly agree, 1 = strongly disagree, as in the examples above.

References

- Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.
- Ambridge, B. (2011). Paradigms for assessing children's knowledge of syntax and morphology. In E. Hoff (Ed.), *Guide to research methods in child language*. London: Blackwell-Wiley.
- Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (in press). The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language*.
- Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, 106(1), 87–129.
- Ambridge, B., Pine, J. M., Rowland, C. F., Jones, R. L., & Clark, V. (2009). A semantics-based approach to the "no negative evidence" problem. *Cognitive Science*, 33(7), 1301–1316.
- Ambridge, B., Pine, J. M., & Rowland, C. F. (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics*, 22(2), 303–323.
- Anderson, S. R. (1971). On the role of deep structure in semantic interpretation. *Foundations of Language*, 6, 197–219.
- Baayen, H. R. (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 Enquiry*, 10, 533–581.
- Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Bavin, E. L., & Growcott, C. (2000). Infants of 24–30 months understand verb frames. In S. Perkins & S. Howard (Eds.), *New directions in language development and disorders* (pp. 169–177). New York: Kluwer.
- Beavers, J. (2010). The structure of lexical meaning: Why semantics really matters. *Language*, 86, 821–864.
- Berwick, R. C., & Weinberg, A. S. (1984). *The grammatical basis of linguistic performance: Language use and acquisition*. Cambridge, MA: MIT.
- Boland, J. E., & Tanenhaus, M. K. (1991). The role of lexical representations in sentence processing. In G. Simpson (Ed.), *Understanding word and sentence* (pp. 331–366). New York: North-Holland.
- Bowerman, M. (1981). *The child's expression of meaning: Expanding relationships among lexicon, syntax and morphology*. Paper presented at the New York academy of sciences conference on native language and foreign language acquisition.
- Bowerman, M. (1988). The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford: Blackwell.
- Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, 85, 55–83.
- Boyd, J. K., Gottschalk, E. A., & Goldberg, A. E. (2009). Linking rule acquisition in novel phrasal constructions. *Language Learning*, 59, 64–89.
- Braine, M. D. S., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 352–376). Hillsdale, NJ: Erlbaum.
- Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, 75(4), 720–738.
- Brooks, P. J., & Zizak, O. (2002). Does preemption help children learn verb transitivity. *Journal of Child Language*, 29, 759–781.
- Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development*, 70(6), 1325–1337.
- Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, 8(6), 500–508.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30(3), 637–669.
- Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 23, 696–708.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31(1), 41–81.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9, 427–473.
- Ford, M., Bresnan, J., & Kaplan, R. M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.
- Gahl, S. (2002). The role of lexical biases in aphasic sentence comprehension. *Aphasiology*, 16, 1173–1198.
- Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, and Computers* (36), 432–443.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language, thought, and culture* (Vol. 2, pp. 301–334). Hillsdale, NJ: Erlbaum.
- Gentner, D. (1975). Evidence for the psychological reality of semantic components: The verbs of possession. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 211–246). San Francisco: W.H. Freeman.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, 49, 988–998.
- Gleitmann, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3–55.
- Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, 22, 131–154.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gropen, J. S., Hollander, M., & Goldberg, R. (1991a). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, 41, 1–3.
- Gropen, J. S., Hollander, M., & Goldberg, R. (1991b). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language*, 18, 115–151.
- Hirsh-Pasek, K., Golinkoff, R. M., & Naigles, L. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT.
- Iwata, S. (2008). *Locative alternation: A lexical-constructional approach*. Amsterdam: John Benjamins.
- Jeffries, L., & Willis, P. (1984). A return to the spray/paint issue. *Journal of Pragmatics*, 8, 715–729.
- Jennings, F., Randall, B., & Tyler, L. K. (1997). Graded effects of verb subcategory preferences on parsing: Support for constraint-satisfaction models. *Language and Cognitive Processes*, 12(4), 485–504.
- Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The What's X doing Y? Construction. *Language*, 75, 1.
- Kidd, E., Bavin, E., & Rhodes, B. (2001). *Two-year-olds' knowledge of verbs and argument structures* (pp. 1368–1382). Cascadilla Press.
- Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63). Stanford: CSLI.
- Lapata, M., Keller, F., & Walde, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, 30, 419–435.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

- Levin, B., & Rappaport Hovav, M. (1991). Wiping the slate clean: A lexical semantic exploration. *Cognition*, 41, 123–151.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157–201.
- MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31(4), 883–914.
- Naigles, L. G., & Kako, E. T. (1993). 1st Contact in verb acquisition – Defining a role for syntax. *Child Development*, 64(6), 1665–1687.
- Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58(2), 221–251.
- Naigles, L., Fowler, A., & Helm, A. (1992). Developmental shifts in the construction of verb meanings. *Cognitive Development*, 7, 403–427.
- Naigles, L., Gleitman, H., & Gleitman, L. R. (1993). Syntactic bootstrapping and verb acquisition. In E. Dromi (Ed.), *Language and cognition: A developmental perspective*. NJ: Ablex.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT.
- Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, 92, 377–410.
- Rappaport Hovav, M., & Levin, B. (1985). *A case study in lexical analysis: The locative alternation*. Unpublished manuscript, MIT Centre for Cognitive Science.
- Stefanowitsch, A. (2008). Negative evidence and preemption: A constructional approach to ungrammaticality. *Cognitive Linguistics*, 19(3), 513–531.
- Stefanowitsch, A. (2011). Constructional preemption by contextual mismatch: A corpus-linguistic investigation. *Cognitive Linguistics*, 22, 107–129.
- Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgement tasks. *Cognitive Development*, 19(1), 15–34.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–553.
- Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.
- 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. A., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56, 165–209.
- Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (submitted for publication). Novel construction learning in 5 year olds: The role of input and age.