ResearchGate

See discussions, stats, and author profiles for this publication at:
https://www.researchgate.net/publication/233387245

Avoiding dative overgeneralisation errors: Semantics, statistics or both?

ARTICLE · SEPTEMBER 2014

DOI: 10.1080/01690965.2012.738300

CITATIONS
12

READS
133

5 AUTHORS, INCLUDING:

Julian Mark Pine
University of Liverpool
109 PUBLICATIONS **2,813** CITATIONS

SEE PROFILE

Caroline F Rowland
University of Liverpool
59 PUBLICATIONS **1,367** CITATIONS

SEE PROFILE

Available from: Ben Ambridge
Retrieved on: 09 April 2016

LANGUAGE AND COGNITIVE PROCESSES
0000, 00 (00), 1–37

Psychology Press
Taylor & Francis Group

Avoiding dative overgeneralisation errors: Semantics, statistics or both?

Ben Ambridge[1], Julian M. Pine[1], Caroline F. Rowland[1], Daniel Freudenthal[2], and Franklin Chang[1]

AQ1 [1]University of Liverpool, Liverpool, UK
[2]Department of Health and Society, Institute of Psychology, University of
AQ2 Liverpool, Liverpool, UK

How do children eventually come to avoid the production of overgeneralisation errors, in particular, those involving the dative (e.g., **I said her "no"*)? The present study addressed this question by obtaining from adults and children (5–6, 9–10 years) judgements of well-formed and over-general datives with 301 different verbs (44 for children). A significant effect of *pre-emption*—whereby the use of a verb in the prepositional-object (PO)-dative construction constitutes evidence that double-object (DO)-dative uses are not permitted—was observed for every age group. A significant effect of entrenchment—whereby the use of a verb in any construction constitutes evidence that unattested dative uses are not permitted—was also observed for every age group, with both predictors also accounting for developmental change between ages 5–6 and 9–10 years. Adults demonstrated knowledge of a morphophonological constraint that prohibits Latinate verbs from appearing in the DO-dative construction (e.g., **I suggested her the trip*). Verbs' semantic properties (supplied by independent adult raters) explained additional variance for all groups and developmentally, with the relative influence of narrow- vs broad-range semantic properties increasing with age. We conclude by outlining an account of the formation and restriction of argument-structure generalisations designed to accommodate these findings.

Keywords: Child language acquisition; Retreat from overgeneralisation; Entrenchment; Pre-emption; Semantic verb class hypothesis.

A central question in psycholinguistics is that of how children attain adult competence in their native language. The hallmark of adult linguistic competence is the ability to produce utterances that are entirely novel, yet they are regarded as grammatically acceptable by fellow native speakers. Thus, an account of how children acquire this ability will form the core of any successful theory of language acquisition.

Correspondence should be addressed to Ben Ambridge, School of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford St South, Liverpool L69 7ZA, UK. E-mail: Ben.Ambridge@Liverpool.ac.uk

This research was supported by grants from the Economic and Social Research Council (RES-062-23-0931) and the Leverhulme Trust (RPG-158).

© 2012 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2012.738300

UNCORRECTED PROOF

This problem is complicated by what is sometimes known as "Baker's paradox" (Baker, 1979). On the one hand, children must form generalisations that allow them to extend verbs that have been attested in one particular construction into nonattested constructions. On the other hand, children must somehow avoid extending certain verbs into nonattested constructions, in order to avoid producing utterances that would be viewed as ungrammatical by adult speakers. For example, many verbs that appear in the prepositional-object (henceforth PO) dative construction (e.g., *Bill gave a present to Sue*) may also appear in the double-object (henceforth DO) dative construction (e.g., *Bill gave Sue a present*). The formation of this generalisation, therefore, allows children to produce novel utterances. For example, a child who heard *John texted the directions to Mary* (PO) could produce an utterance such as *John texted Mary the directions* (DO) without having ever encountered the verb *text* in the DO-dative construction. However, children must somehow avoid applying this generalisation to verbs for which it would yield an ungrammatical "overgeneralisation error" (e.g., *I said "no" to her* → **I said her "no"*).[1] Although overgeneralisation errors of exactly this type have been observed in studies of children's spontaneous speech (e.g., Bowerman, 1988; Pinker, 1989), Baker's paradox applies equally to children who do and do not produce such errors. The question is how children either retreat from such errors or avoid them altogether, whilst retaining the capacity for productivity that characterises mature linguistic competence.[2]

Current attempts to answer this question can be divided into three basic types: semantics-based, statistics-based, and hybrid-based approaches. The aim of the present study is to attempt to clarify precisely the type of semantics-based and statistics-based learning mechanisms that are needed to account for participants' grammaticality judgement data, and how best to integrate them into an account of learning and development.

[1] A reviewer asked when children typically acquire the PO-/DO-dative distinction. The age at which children acquire the subtly different semantic properties of the two constructions is one of the questions investigated in the present study. But at what age are children able to correctly use and understand the two constructions in a more general sense (i.e., understand that the PO-dative places the theme before the recipient, whilst the DO-dative displays the opposite pattern)? Studies of spontaneous speech (e.g., Campbell & Tomasello, 2001; Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989; Snyder & Stromswold, 1997) suggest that both emerge in production at around 2;0. There is some suggestion that the DO-dative appears earlier, though this likely largely reflects the use of semi-productive chunks such as *Gimme X*. A recent forced-choice comprehension study (Rowland & Noble, 2010) found that 3-4 year-olds demonstrated more robust performance with PO- than DO-datives, but could still correctly interpret DO-dative sentences when given additional cues (e.g., when the recipient was a proper noun).

[2] Baker's paradox is also sometimes known as the 'no negative evidence' problem (e.g., Bowerman, 1988; Marcus, 1993), as the assumption is that caregivers do not provide children with evidence regarding the (un)grammaticality of their utterances (e.g., McNeill, 1966). In fact, this is something of a misnomer, as evidence suggests not only that caregivers spontaneously correct such errors, but also that children are sensitive to this feedback (e.g., Chouinard & Clark, 2003; Clark & Bernicot, 2008; Saxton, Backley & Gallaway, 2005; Strapp, Bleakney, Helmick, & Tonkovich, 2008). However, this type of feedback is unlikely to be sufficient as an explanation of the acquisition of restricted generalizations. One problem is that both children and adults rate as ungrammatical certain uses of low frequency and even novel verbs (e.g., Ambridge, Pine, Rowland, & Young, 2008; Wonnacott, Newport & Tanenhaus, 2008), for which they cannot have received such feedback. A second problem is that such errors - and hence opportunities for feedback - are relatively rare; indeed some children may not produce them at all. Yet Baker's paradox applies equally to children who do not produce such errors, where the question becomes how they avoid them whilst retaining the ability to produce novel utterances.

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:16

SEMANTICS-BASED APPROACHES

Verbs that may and may not appear in particular constructions do not form wholly arbitrary lists, but generally share certain semantic properties (and, in some cases, properties related to other factors such as morphophonology or pragmatics). For example, it has often been observed that verbs that are more felicitous in the PO- than DO-dative are associated with the meaning of *the THEME being caused to go to a LOCATION, often in a particular manner.*

[AGENT] [VERB] [THEME] to [LOCATION/RECIPIENT]
John lowered the package to Sue. (cf. *?John lowered Sue the package)
John pulled the box to Sue. (cf. *?John pulled Sue the box)
John whispered the secret to Sue (cf. *?John whispered Sue the secret)

Conversely, verbs that are more felicitous in the DO- than PO-dative are associated with the meaning of **the RECIPIENT being caused to possess (or to no longer possess) the THEME**:

[AGENT] [VERB] [RECIPIENT] [THEME]
John tipped Sue $5 (cf. *?John tipped $5 to Sue)
John's advice saved Sue a lot of money (cf. *?John's advice saved a lot of money to Sue)
John cost Sue a fortune (cf. *?John cost a fortune to Sue).

Verbs that alternate between the two constructions (e.g., *give, send, pass*) do so because they have elements of both *causing to go* and *causing to have* in their meanings. Corpus studies suggest that these are not hard-and-fast prohibitions, but a graded phenomenon whereby the use of a verb in a construction with which it is less than optimally compatible is deemed to be less than fully acceptable (e.g., Bresnan, Cueni, Nikitina, & Baayen, 2007).

The claim of semantics-based approaches is that learners restrict their generalisations by acquiring the semantic restrictions exhibited by particular constructions. A particularly well-specified proposal for how this might be done is provided by Pinker (1989). Pinker proposed that to be a candidate for appearing in the PO- and DO-dative, a verb must be consistent with the broad-range "thematic core" of each construction: *causing to go* and *causing to have*, respectively. This is a necessary, but not sufficient, criterion. Actual appearance in each construction is contingent on membership of a relevant narrow-range semantic class. For example, *give* and *hand* are members of a class of verbs ("verbs of giving"; p. 110) that appear in the DO-dative construction (e.g., *I gave/handed him the box*). *Lower* and *pull* are not members of such a class (they are members of the PO-only class of "accompanied motion"), and hence may not appear in this construction, notwithstanding that they are potentially compatible with the broad-range rule for the DO-dative (possession transfer). Children form classes—and hence constrain their generalisations—by extending observed properties of particular verbs (e.g., that *give* appears in the DO-dative construction) to "other verbs with the same grammatically relevant semantic structure" (e.g., *send*, *pass*).

Both the broad-range rules and the narrow-range classes have proved controversial. The broad-range rule, whereby the PO-dative denotes *causing to go* and the DO-dative *causing to have* is often invoked to explain contrasts such as:

PO: John sent the package to Chicago vs DO: *John sent Chicago the package
DO: The noise gave Paul a headache vs PO: *The noise gave a headache to Paul.

UNCORRECTED PROOF

In the first example, the package is caused to go to Chicago; Chicago is not caused to possess the package. Hence only the PO-variant is grammatical. In the second example, Paul is caused to possess a headache; the headache is not transferred from the noise to Paul. Hence only the DO-variant is grammatical. However, a number of recent studies have found that such "violations" are relatively frequent in large corpora and that—for both adults and children—dative choice in production is determined largely by factors such as the relative animacy, accessibility, definiteness, and length of the recipient and theme (e.g., Bresnan, 2007a, 2007b; Bresnan & Nikitina, 2007; Bresnan et al., 2007; de Marneffe, Grimma, Arnon, Kirby, & Bresnan, 2011; Krifka, 2004). For example, the grammaticality of a sentence of the form **The noise gave a headache to Paul* is much ameliorated if the recipient is a long and complex noun phrase, as such NPs are generally placed sentence-finally for processing reasons (e.g., **?The noise gave a headache to anyone who was unfortunate enough to be in the room at the time*). This raises the possibility that the broad-range semantic constraints proposed by Pinker (1989) may be of relatively minor importance, or even entirely epiphenomenal, arising entirely from these discourse and processing factors.

There has also been considerable scepticism in the literature with regard to Pinker's (1989) narrow-range classes. For example, Braine and Brooks (1995) argue that it is implausible to suppose that children are sensitive to such fine-grained distinctions as that between (for example), *tell*, which may appear in the DO-dative construction, and *say*, which may not (e.g., *He told/*said me something funny*; cf. *He told/said something funny to me*; for Pinker, the distinction is that only the former implies successful transfer of information). Bowerman (1988) argues that it is not clear what would cause children to form and continually refine these narrow-range classes, given that the broad-range rules allow them to understand and produce the relevant utterances (of course, a central part of Baker's paradox is that children do not know if or when they are producing errors).

One recent grammaticality judgement study provides support for Pinker's (1989) semantic verb class hypothesis with respect to the dative constructions, but only for adults. In this study (Ambridge, Pine, Rowland, & Chang, 2012), adults and children aged 5- to 6-year-olds and 9- to 10-year-olds were taught novel verbs consistent with PO-only semantic classes (e.g., similar in meaning to *pull/drag*) and alternating semantic classes (e.g., similar in meaning to *give/send*). Adults respected the semantic class of the novel verbs, rating PO-datives as more acceptable than DO-datives for the novel *pull/drag* verbs (e.g., *Marge tamed the box to Homer* > *Marge tamed Homer the box*), but not the novel *give/send* verbs (e.g., *Bart blicked the package to Marge* = *Bart blicked Marge the package*). Neither of the child groups showed this pattern, rating PO- and DO-dative sentences as approximately equally acceptable for all novel verbs. However, this may be due in part to the fact that the semantic distinction between PO-only and alternating verb classes is relatively fine-grained. Similar studies involving overgeneralisation errors into the transitive-causative construction (e.g., **The funny clown giggled Lisa*) have found that both 9- to 10-year-olds and 5- to 6-year-olds display at least some semantic-class effects with novel verbs (Ambridge, Pine, & Rowland, 2011; Ambridge et al., 2008).

The first aim of the present study was to investigate in more detail the effects of verb semantics on dative overgeneralisation errors. The first question of interest is precisely which of the many different semantic features exhibited by PO-only, DO-only, and alternating verbs learners are sensitive to, and whether this changes with development. The second question is whether whatever sensitivity learners show to these semantic features is best captured by positing broad-range rules and discrete

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:17

semantic verb classes (as under Pinker's account), or whether a more probabilistic learning mechanism is required. One reason to consider the latter possibility is the finding that verb frequency effects are pervasive in the literature on the retreat from overgeneralisation errors, and it is debatable whether Pinker's account, at least in its present form, can explain such effects. It is to proposals seeking to explain these effects that we now turn.

STATISTICS-BASED APPROACHES

The simplest statistics-based approach is the *entrenchment* hypothesis (e.g., Ambridge et al., 2008; Braine & Brooks, 1995; Brooks, Tomasello, Dodson, & Lewis, 1999; Theakston, 2004; Tomasello, 2003). Although, slightly more advanced versions have been proposed (e.g., Stefanowitsch, 2008), the most commonly assumed mechanism requires learners simply to tally in the input language (1) the total number of occurrences of a particular verb (e.g., *say*) and (2) the number of occurrences of this verb in the target construction (e.g., the DO-dative). Provided that (2) remains essentially zero (i.e., the occasional speech error notwithstanding), each additional count of (1) contributes to a probabilistic inference that the use of the verb in this construction is ungrammatical (e.g., **I said her no*). This is a classic "inference from AQ36 absence" (Hahn & Oaksford, 2008) of the type that is captured by Bayesian rational-learning models (e.g., Alishahi & Stevenson, 2008; Chater & Vitanyi, 2007; Dowman, 2000; Hsu, 2009; Hsu & Chater, 2010; Onnis, Roberts, & Chater, 2002; Perfors, Tenenbaum, & Wonnacott, 2010).

The *pre-emption* hypothesis (e.g., Boyd & Goldberg, 2011; Clark & Clark, 1979; Goldberg, 1995Goldberg, 2011) is similar to entrenchment, but with one important difference. Under this account, the use of a verb in a particular construction contributes to the inference that a particular nonattested usage is ungrammatical only if this target usage was at least as felicitous, given the discourse context and functional demands of the situation.[3] For example, all other things being equal, a PO-dative use of *say* (e.g., *I said "no" to her*) contributes to the inference that the DO-dative equivalent (**I said her "no"*) is ungrammatical, because this target formulation is at least as felicitous, given the speaker's intended message. Occurrences of *say* in other constructions (e.g., the simple transitive as in *I said "no"*) are irrelevant, because they are not competing with the target utterance to express (virtually) the same meaning. This contrasts with entrenchment, under which hearing a verb in *any* construction contributes to the inference that nonattested uses are not permitted.

Dissociating the mechanisms of entrenchment and pre-emption is important, given the wider debate about which aspects of language can be learned on a purely statistical basis (e.g., Saffran, Aslin, & Newport, 1996) and which also require meaning and discourse function to be taken into consideration. However, it is extremely difficult to

[3] In the corpus analysis conducted as part of the present study, we make the simplifying assumption that any occurrence of a PO-dative constitutes an instance of case where a DO-dative would have been equally felicitous (and vice versa). This is an oversimplification as factors such as the relative animacy, accessibility, definiteness, and length of the recipient and theme affect the relative felicitousness of the two constructions (see Bresnan et al., 2007). However, hand-coding each dative sentence on each of these dimensions would have been not only prohibitively time-consuming, but also—for some dimensions—difficult to do objectively (though see Goldberg, 2011, for a study along these lines). Consequently, the present study constitutes a particularly strong test of the pre-emption hypothesis, because the noisiness of the pre-emption measure that results from this simplification counts against the likelihood of any such effect being observed.

UNCORRECTED PROOF

do so with corpus data, since the two predictors are inevitably highly correlated ($r = 0.9$ for the present dataset) (though see Goldberg, 2011; Stefanowitsch, 2008, 2011). Thus whilst, in the present study, we attempt to do so as best we can, we acknowledge that dissociating these two mechanisms will likely require direct experimental manipulation. Note that most existing experimental studies that are described as testing either entrenchment (e.g., Ambridge, Pine, Rowland, Jones, & Clark, 2009; **AQ5**,**AQ4** Ambridge et al., 2008, 2011, 2012; Brooks et al., 1999; Perfors et al., 2010; Theakston, 2004; Wonnacott, 2011; Wonnacott et al., 2008) or pre-emption (Brooks & Tomasello, 1999) do not in fact attempt to dissociate between the two proposals (exception are Boyd & Goldberg, 2011; Brooks & Zizak, 2002).

SEMANTICS VS STATISTICS

The third aim of the present study is to test different proposals regarding the relative contributions of semantics and statistics. The main positions that have been taken on this issue are as follows.

Semantics only

Under Pinker's (1989) semantic verb class hypothesis, the formation of narrow-range semantic verb classes is the sole mechanism by which children retreat from error. Of course, frequency effects are not necessarily incompatible with this approach, in principle. For example, one can imagine a version of this theory under which low-frequency verbs take longer to be assigned to the correct semantic class (though Pinker himself does not discuss this possibility). However, the theory as it currently stands includes no role for any type of statistical inference-from-absence (i.e., entrenchment/pre-emption). A different type of semantics-only proposal was outlined by Ambridge et al. (2009), who suggested that children retreat from error by refining their knowledge of the semantic properties of particular verbs and constructions, to the point where they can detect any mismatch between the two (held to be the source of the ungrammaticality of some generalisations). Frequency effects were argued to be entirely epiphenomenal, arising from the fact that more frequent verbs simply have better-learned semantics.

Statistics only

This position is exemplified by Stefanowitsch (2008, p. 527), who argues 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". Under this view, semantic considerations might explain why the language is the way that it is, but children learn verbs' argument-structure privileges on a purely distributional basis, via entrenchment or pre-emption.

Semantics with statistics for exceptions

A closely related position is that whilst certain argument-structure restrictions have a semantic basis, which children may identify and make use of in learning, some restrictions are purely arbitrary and can be learned only by entrenchment and/or pre-emption. This position is exemplified by Boyd and Goldberg (2011 p. 58), who argue that:

UNCORRECTED PROOF

> while motivation for restrictions is often available, we need to keep in mind that this type AQ33 of explanation does not predict the illformedness of [certain] examples.... This is evidenced by the fact that expressions that share closely related semantics, pragmatics and phonological properties are fully acceptable.

It is almost certainly true that some verbs have some entirely idiosyncratic selection properties (e.g., *manage to do*; *succeed in doing* vs **succeed to do*; **manage in doing*). The position exemplified by Boyd and Goldberg (2011) treats more abstract constructions in the same way. This partly motivated, partly idiosyncratic view is probably the most widely held in the literature. For example, both Bowerman (1988) and Braine and Brooks (1995, p. 366), whilst acknowledging some role for verb semantics, argue that Pinker's classes are "riddled with exceptions".

Statistics then semantics

Tomasello (2003, p. 180) argues that "entrenchment works early ... and semantic subclasses begin to work later, perhaps not until about 4;6 or so". Similarly, Perfors, AQ6 Tenenbaum, and Regier (2011, p. 636) argue that their study "suggests the possibility that although semantic information is ultimately used, syntactic information may be more important initially". Wonnacott (2011, p. 14), whilst appearing to advocate "at least some arbitrary lexical specification" (see above), notes that: "it is also clear that adults and *older* children [emphasis added] are sensitive to semantic and phonological regularities". A number of findings provide support for this view. For example, Brooks AQ34 and Tomasello (1999) found that children aged 4;5 and above produced more transitive-causative utterances for novel verbs from an alternating semantic class (e.g., *The mouse tamed [=span] the ball*) than an intransitive-only class (e.g., **The mouse meeked [=ascended] the ball*), but a younger group aged 2;5 showed no such effect. However, in a similar study (Brooks et al., 1999), children aged 3;4 and upwards showed frequency (i.e., entrenchment/pre-emption) effects with real English verbs (e.g., **The funny clown laughed/giggled the mouse*). Similarly, in the dative study of AQ7 Ambridge et al. (2012), all age groups displayed an effect of entrenchment/pre-emption, with only the adults displaying an effect of semantic verb classes (see also Ambridge et al., 2008, for transitive-causative errors).

A problem with these studies, however, is that they do not compare like with like. In order to display a semantic effect, children must demonstrate a particular pattern of production/judgements with *novel* verbs. In contrast, children can display an entrenchment/pre-emption effect simply by showing different performance with two *familiar* verbs (e.g., **The funny clown laughed/giggled Lisa*). This is a fundamental problem because any apparent finding that entrenchment/pre-emption emerges before effects of verb semantics may, in fact, simply reflect increased task difficulties associated with novel verbs. Indeed, in a study using only novel verbs whose semantics and frequency were manipulated experimentally, children aged 5—6 years displayed an effect of semantics but not frequency (Ambridge et al., 2011). This is not to argue that semantics are *more* important than frequency early in development, but simply to illustrate that the relative likelihood of observing each effect varies as a function of the task. In the present study, we aim to achieve as level a playing field as possible by comparing the effects of entrenchment/pre-emption and verb semantics on a large set of familiar verbs.

To our knowledge, only one previous study has attempted to compare the relative contributions of semantic and statistical factors in this way. Ambridge, Pine, and Rowland (2012) applied essentially the same methodology used in the present study to

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:17

overgeneralisation errors involving the locative alternation (e.g., *Lisa poured water into the cup*; **Lisa poured the cup with water*; **Bart filled water into the cup*; *Bart filled the cup with water*). For every age group, the optimal statistical model included both semantic and statistical predictors. With regard to semantics, the influence of both the broad- and narrow-range rules increased with age, and roughly in parallel. With regard to statistics, although both entrenchment and pre-emption were significant predictors of participants' judgements independently, entrenchment was the only one of the two predictors to explain additional variance above and beyond that explained by the other (although the two predictors were highly correlated, $r = 0.70$, $p < .001$, making it extremely difficult to estimate the independent contributions of each). On the assumption that essentially the same learning process occurs for all types of verb argument-structure restrictions, the expectation is that a very similar pattern will be observed in the present study.

Summary

The present study investigates how children retreat from, or avoid, argument-structure overgeneralisation errors involving the dative constructions (e.g., **I said her "no"*). In particular, the study has two aims. The first is to investigate whether adults and children are sensitive to the types of broad- and narrow-range semantic verb properties proposed by Pinker (1989) and, if so, whether this is best captured by positing discrete verb classes or a more probabilistic learning mechanism. The second is to compare different proposals regarding the relative contributions of semantics and statistics at different points in development (*semantics only*; *statistics only*; *semantics, with statistics for exceptions*; *statistics then semantics*).

In order to allow for investigation of these various factors, we obtained—for each of 301 dative verbs—measures of entrenchment and pre-emption (from the British National Corpus) and measures of the extent to which each verb instantiates particular broad- and narrow-range semantic properties (from independent adult raters). We then assessed the ability of these different predictors to account for judgements of PO- and DO-dative uses of (for adults) all 301 verbs or (for children) a representative set of 44 verbs.

Before proceeding, a note is in order regarding our choice of paradigm. The phenomenon under investigation is that (some) children make overgeneralisation errors in production, which has generally been taken as evidence of incomplete underlying grammatical knowledge. Our method assumes that grammaticality judgements tap into this underlying knowledge (of course, somewhat indirectly), and share some meaningful relationship with production (i.e., that children who judge particular errors as relatively acceptable would produce them, given an appropriate context). This assumption has often been challenged on *a priori* theoretical grounds. For example, Edelman and Christiansen (2003) argue that judgement tasks assess metalinguistic knowledge of language, rather than language as it is actually used by speakers. Empirically, however, as Gibson and Fedorenko (2011, p. 5) point out "results from acceptability judgement experiments are highly systematic across speakers and correlate with other dependent measures, presumably because the same factors affect participants' responses across different measures" (Schutze, 1996; Sorace & Keller, 2005). Indeed, the goal of the present study is to elicit judgements of exactly the type of errors that have been observed in spontaneous production data (e.g., Bowerman, 1988). Whilst we acknowledge that the judgement paradigm does not constitute a clear window onto children's underlying knowledge, it is not clear that

UNCORRECTED PROOF

there exists any paradigm that provides a more direct measure. For example, in even apparently straightforward production tasks, children's responses are heavily affected by the particular prompts used by the experimenter (e.g., Arnon & Clark, 2011).

That said, it is important to be mindful of the potential confounds introduced by the judgement task, and, in particular, the possibility that children may approach the task in a qualitatively different way to adults. In the present study, we attempt to minimise the effect of such confounds by tightly controlling the stimuli (e.g., each DO-/PO-sentence pair uses the same NPs), normalising the data on a participant-by-participant basis (i.e., using *z*-scores), and confirming that children's judgements are relatively adult-like in cases where we would expect them to be so (e.g., when rating unambiguously grammatical sentences).

METHOD

Participants

Participants were 36 children aged 5;2–6;1 ($M = 5;7$), 36 children aged 9;2–10;1 ($M = 9;8$), and 30 adults aged 18–21, with an equal number of males and females at each age. A further 15 adults were recruited to provide ratings of verbs' semantic (and morphophonological) properties, as detailed below. All participants were normally developing monolingual speakers of British English, and were primarily from a **AQ8** middle-class background (though detailed SES information was not collected). Children were tested at their school in the North West of England. The adults were first year undergraduate Psychology students. None had studied language acquisition or linguistics.

Materials and procedure

Items

The study included all 301 verbs listed as dative by either Pinker (1989) or Levin (1993), subsequently referred to as the extended set. According to the classification of these authors, this comprised 145 alternating verbs, 131 PO-only verbs and 25 DO-only verbs.[4] However, it is important to note that these classifications were not used in the statistical analysis (with one exception); rather, objective ratings of the extent to which verbs exhibit particular semantic properties relevant to the class definitions were used. For each verb, adults rated the acceptability of one PO-dative and one DO-dative sentence using a written questionnaire. Two complete sets of PO-/DO-sentence pairs—1 for each of the 301 verbs—were created. Within a given set, each PO-/DO-pair used the same NPs, but different NPs to those used in the corresponding pair in the other set. For example, considering the verb *give*, half of the participants rated both *Homer gave a book to Bart* (PO) and *Homer gave Bart a book* (DO), whilst the other half rated both *The shopkeeper gave a bag to the customer* (PO) and *The shopkeeper gave the customer a bag* (DO). Within each set, three different pseudo-random presentation orders were used.

[4] For the purposes of these descriptive counts, a verb was classified as PO-only or DO-only if either or both of the two authors listed it as such, otherwise it was classified as alternating. Agreement between the two sources was high with only seven verbs classified as alternating by Levin and PO-only by Pinker (*carry, haul, pull, push, schlep, repeat, recount*) and two as DO-only by Levin and alternating by Pinker (*guarantee, declare*). No verb was classified as PO-only by one author and DO-only by the other.

UNCORRECTED PROOF

UNCORRECTED PROOF

Forty-four verbs (subsequently referred to as the core set) were selected for use with children, divided equally between PO-only and alternating verbs, as classified by Pinker/Levin (though it is important to remember that these classifications are descriptive only, and play no role in the main statistical analyses):

> Alternating: feed, give, pass, sell, post, send, hit, kick, throw, toss, bring, take, award, offer, owe, promise, read, show, teach, tell, email, telephone.
>
> PO-Only: carry, haul, pull, push, drag, drop, heave, hoist, lift, lower, raise, mumble, scream, shout, whisper, yell, mention, say, state, entrust, present, supply.

Care was taken to ensure that all the semantic classes listed by Pinker (1989) and Levin (1993) were represented in this selection. Although this entailed the inclusion of some verbs that are likely to be unfamiliar to children (particularly the younger group), this was considered preferable to leaving some classes unrepresented. Furthermore, the inclusion of some unfamiliar verbs could even be seen as desirable, as Pinker's (1989) semantic verb class hypothesis predicts that children will make errors precisely because they have yet to acquire the exact meanings of some verbs (some broad sense of each verb's meaning should be inferable from the syntactic context and the accompanying animation). That said, we did not include DO-only verbs—e.g., *bet*, *wager*—as these are likely to be conceptually unfamiliar to young children. They also represent something of a marginal phenomenon, making up only 8% of the verbs rated by adults.

The 44 PO-/DO-sentence pairs were split into two sets of 22 pairs, with each child rating both the PO- and DO-sentence in each pair. As for adults, three different pseudo-random orders were used within each set. To make the study more engaging for children, each sentence was spoken by a talking dog puppet, who the children were "helping to learn English", and accompanied by a cartoon animation presented on a laptop computer.

Rating scale

Adults rated the acceptability of sentences using a 7-point Likert scale. Children used a 5-point "smiley face" scale (see Figure 1). Details of this procedure can be found in Ambridge et al. (2008) and Ambridge (2011). In brief, children indicate the perceived degree of (un)acceptability by placing a red or green counter on a particular face. Nine training sentences containing correct past-tense forms and overgeneralisation errors (e.g., *Lisa ate/*eated the ice cream*) were used to train participants in the appropriate use of the scale.

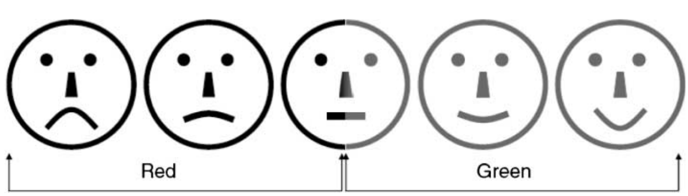

Figure 1. Scale used by children to rate the acceptability of sentences (reproduced from Ambridge et al., 2008, p. 105).

UNCORRECTED PROOF

Statistical predictors

Two corpora were used to derive the statistical predictors: the (adult–adult) British National Corpus (subdivided into all texts and spoken only) and the child-directed-speech portion of the post-Manchester Corpus (Rowland & Theakston, 2009). As both corpora are tagged for part-of-speech categories, the counts for the entrenchment predictor—the raw number of VERB uses of each verb—were extracted automatically (then log $N+1$ transformed).[5] Two additional modifications were necessary. The entrenchment hypothesis predicts that the higher the frequency of a particular verb in all attested constructions, the stronger the inference that the use of this verb in a particular target construction is not allowed, *unless it has been independently attested*. Therefore, for any verb attested in both the PO- and DO-dative constructions, frequency was set to zero. Second, the sign of the log-transformed entrenchment predictor was set to positive for verbs that appear in the corpus in PO-datives only, and negative for verbs that appear in the corpus in DO-datives only. This is because the outcome variable (difference score) is positive for verbs that participants consider to be more acceptable in the PO- than DO-dative and negative for verbs that display the opposite pattern.

The pre-emption measure was calculated in exactly the same way, except that the counts were based on the number of occurrences of each verb *in the relevant dative construction* as opposed to overall. For any given corpus, the entrenchment and pre-emption predictors are inevitably highly correlated, as the latter counts are a **AQ9** subset of the former (for the present extended set in the BNC, the correlation is $r = 0.90$, $p < .001$). In order to allow us to investigate whether entrenchment and pre-emption explain independent variance, we created an entrenchment measure residualised against pre-emption, and vice versa. Because neither the BNC nor the post-Manchester corpus is syntactically parsed (and the largest parsed corpus, **AQ10** ICE-GB, is too small for our purposes, with 227 of the 301 verbs unattested, as opposed to just 10 in the BNC) counts were extracted using a computer program custom written by the final author (see Section A of the online supplementary material for details).

Semantic predictors

A separate group of 15 adults rated the extent to which they considered each verb to exhibit each of 18 semantic predictors relevant to Pinker's (1989, p. 82) broad-range rules for the 2 constructions and each of 14 semantic predictor relevant to Pinker's (1989, pp. 110–113) narrow-range classes (see Section B of the online supplementary material for details for details). A written questionnaire was used, with the verbs and properties presented in a different pseudo-random order for each participant. Participants provided their ratings using a 9-point numerical scale, and were paid £50.

The aim of collecting these ratings was not to test the psychological validity of the particular broad- and narrow-range classes proposed by Pinker (1989). As Pinker points out, the make-up of individual classes is of no particular theoretical importance, and the classifications posited are not intended to be exhaustive or definitive. Rather, the aim was to investigate whether participants' grammaticality judgements are sensitive to the types of feature posited under the semantic verb class

[5]The log transformation of frequency counts is standard practice in psycholinguistics as these data follow a Zipfian, as opposed to normal, distribution. It is necessary to transform $N + 1$, as opposed to N, as some verbs have a frequency of zero, which has no log.

hypothesis, and—more generally—any account under which semantics plays a key role in restricting argument-structure generalisations. This means that it would not have been appropriate to use narrow-range semantic class membership as a categorical predictor. Furthermore, as an anonymous reviewer noted, to do so would be to introduce an unacceptable degree of circularity, since both Pinker's and Levin's classes were explicitly formed on the basis of verbs' alternation behaviour as well as semantics.

Morphophonological predictors

Following Green (1974) and Oehrle (1976), Pinker (1989) argues that, for at least some semantic classes, verbs that are of Latinate origin are restricted to the PO-dative construction, with DO-datives ungrammatical (e.g., *John suggested the trip to Sue*; **John suggested Sue the trip*). The claim is that when Latinate French words entered the English language, they brought with them their argument-structure ([AGENT] [VERB] [THEME] *à/to* [RECIPIENT]), and resisted assimilation into the argument-structure pattern exhibited by native Germanic verbs such as *give* ([AGENT] [VERB] [RECIPIENT] [THEME]). The claim is not that speakers are sensitive to etymology *per se*. Latinate verbs can be identified by their stress pattern (disyllabic with second-syllable stress [e.g., *donate*] or trisyllabic [*contribute*]), and perhaps by their use of certain morphemes (e.g., -*ify*, -*ate*), whereas native Germanic verbs are mostly monosyllabic (e.g., *give*, *tell*) or have first-syllable stress.

Despite some scepticism in the recent literature (e.g., Felbaum, 2005), there is evidence that this constraint is psychologically real for adult English speakers. AQ11 Ambridge et al. (2012) showed that adults (but not children aged 9–10 or 5–6) rejected as ungrammatical DO-dative uses of novel Latinate verbs (e.g., *Bart orgulated Marge the package*) but not novel native-like verbs (e.g., *Bart naced Marge the package*). Because this previous study found no evidence that children have acquired this constraint, and because the core set of verbs rated by children was necessarily of a limited size, the present study did not systematically investigate the Latinate morphophonological constraint for children. However, in order to investigate, and control for, any such effect in the extended verb set rated by adults, we included a measure of Latinate status. Classifying verbs objectively is not straightforward, as some Latinate verbs have Anglicised stress patterns (e.g., *promise*), and the relative contribution of stress and Latinate-sounding morphemes (e.g., -*ify*, -*ate*) is unclear. We, therefore, asked the adults who completed the semantic rating task to additionally rate each verb for the extent to which:

- The verb is Latinate (a Latinate verb is a verb that comes from Latin, usually via a language such as French or Italian; examples include *investigate*, *conceal*, and *deceive*).
- The verb is formal, "classical" or "learned" (this means that the verb is one that would generally be used in formal written speech, as opposed to everyday casual speech. For example, *conceal* is a relatively formal version that can be used in place of the more everyday verb *hide*).

It was felt that the participants, who had not studied etymology or linguistics, would be unable to provide meaningful ratings in the absence of example verbs (which were not included in the main study).

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:18

RESULTS

The aim of the present study was to investigate the extent to which measures of entrenchment (raw number of verb uses), pre-emption (number of PO- vs DO-verb uses), broad- and narrow-range semantic properties, and morphophonological properties (all obtained from independent adult raters) could predict participants' preference for PO- over DO-dative verb uses. This prediction was tested using a series of linear mixed effects regression (lmer) models in the statistics package *R*.

Before these analyses were conducted, principal components analysis (PCA) was used to condense the 18 broad-range rule predictors (7 relating to the PO-dative; 11 to the DO-dative) into a more manageable number. The Eigenvalue >1 criterion (Kaiser, 1960) was used, resulting in the extraction of a single factor relating to the broad-range rule for the PO-dative (Eigenvalue = 4.11, accounting for 58.67% of total variance) and two factors relating to the broad-range rule for the DO-dative (Eigenvalues = 4.68 and 1.65, accounting for 46.82% and 16.54% of variance, respectively). The same method was used to extract four factors relating to the narrow-range rules, as summarised in Table 1, and a single factor relating to the proposed Latinate restriction on the DO-dative (Eigenvalue = 1.96, accounting for 88.01% of variance).

All analyses were conducted on difference scores, calculated, for each participant, by subtracting the rating for each DO-dative (e.g., *Homer gave Bart a book*) from the rating for its PO equivalent (e.g., *Homer gave a book to Bart*). This controls for any

TABLE 1
Factors relating to Pinker's (1989) narrow-range rules extracted using principle components analysis

Factor label	*Speech*	*Mailing*	*Bequeathing*	*Motion*
Eigenvalue	3.78	2.85	1.71	1.36
% Variance	29.10	21.91	13.16	10.43
Transfer is mediated by a separation in time and space (alternating)	−0.161	**0.816**	−0.313	0.151
A instantaneously imparts force in some manner onto B, causing ballistic motion (alternating)	−0.578	0.070	0.038	**0.497**
A continuously imparts force in some manner onto B, causing accompanied motion (PO-only)	−0.644	0.247	0.194	**0.445**
The verb specifies the direction of motion more than its manner (alternating)	−0.593	0.240	**0.248**	**0.546**
A makes some commitment that C can or cannot have B in the future (DO-only)	0.389	0.473	**0.639**	−0.089
B is something that C deserves, needs, or is worthy of (PO-only)	0.445	**0.724**	**0.264**	−0.112
The verb comes from—or is related to—a noun (PO-only/ alternating, depending on class)	0.035	**0.731**	−0.517	−0.097
C benefits from the action (PO-only)	0.290	0.471	**0.659**	−0.143
A successfully causes C to know (/perceive/apprehend/be aware of) B (an idea or stimulus) (alternating)	**0.660**	0.081	0.081	0.464
The verb specifies the manner of speaking (by A) (PO-only)	**0.668**	−0.488	0.001	0.331
B is usually abstract rather than concrete (alternating)	**0.751**	−0.026	−0.112	0.263
The verb specifies A's attitude with respect to the truth of B (an idea or statement) (PO-only)	**0.737**	−0.206	0.005	0.360
The verb specifies an instrument or means of communication (alternating)	0.482	**0.516**	−0.560	0.130

NOTE: The four questionnaire items with the highest positive loadings on each factor are shown in bold type. Note that the descriptive labels are intended only to capture the flavour of each factor and are necessarily imprecise.

UNCORRECTED PROOF

baseline (dis)preferences participants may have for particular verbs. To test the predictions under investigation, linear mixed effects regression models were fitted to this difference-score rating data. For every model, participant and verb were included as random effects. As detailed below, adding by-participant random slopes had virtually no effect on the *t* values for the fixed effects and did not improve model fit. Hence they are not included in the main analysis presented. To control for differences in the scales used by the adults and children, difference scores were converted into *z*-scores on a participant-by-participant basis. Consequently, the coefficients shown in all results tables are standardised coefficients (i.e., units of standard deviation). The predictor variables were not standardised as it is unclear how frequency counts and semantic ratings can meaningfully be transformed into comparable units.

All verbs ($N = 301$ for adults, $N = 44$ for children)

The first analysis was conducted on all ratings combined. That is, it included data from both adults (for the extended set of 301 verbs) and children (for the core set of 44 verbs). Uneven data sets (caused here by the presence of items rated by adults only) do not constitute a problem for linear mixed effects regression analysis. In order to investigate the independent contributions of (1) broad-range rules, (2) narrow-range semantic classes, (3) the Latinate morphophonological constraint, (4) pre-emption, and (5) entrenchment, we started with a single-predictor model and investigated whether adding each subsequent predictor yielded improved coverage of the data. For the optimal model, we then investigated whether adding (1) by-participant random slopes for each effect (see Barr, Levy, Scheepers, & Tiley, submitted) and (2) interaction terms improved model fit. Models were compared using likelihood ratio tests, where lower log likelihood and Akaike's Information Criterion (AIC) values indicate better model fit. The model comparison procedure is described in detail in Section C of the online supplementary material, with the models shown in Table 2 (left hand columns)

The optimal model (Table 2d) included one of the broad-range semantic predictors (the broad-range rule for the PO-dative), all four of the narrow-range predictors and pre-emption (or entrenchment; both explain largely the same variance, though the pre-emption predictor slightly more; see Section C of the online supplementary material for details). In general, adding random slopes and interactions did not improve the model (see Section C of the online supplementary material for details). The morphophonological predictor did not explain a significant amount of variance in this model, though it did if entered before pre-emption (again, see Section C of the online supplementary material for details), suggesting that it may initially be learned on a verb-by-verb basis.

The first narrow-range predictor, *speech*, was derived primarily from semantic features thought to characterise PO-only classes (e.g., manner of speaking), but also some alternating classes (e.g., A successfully causes C to know B). Thus, since the dependent measure reflects participants' preference for PO- over DO-dative uses, one would expect this predictor to have a significant positive correlation with the outcome measure. This pattern was observed. Similarly, the narrow-range predictor *motion* was derived from alternating classes and from the PO-only class "A continuously imparts force in some manner onto B, causing accompanied motion", and hence was also significantly correlated with participants' preference for PO- over DO-uses. The *bequeathing* predictor was something of a mixed bag, with loadings from both PO-only (e.g., C benefits from the action) and DO-only classes (A makes some commitment that C can or cannot have B in the future). This is perhaps why this

AQ30

TABLE 2
Results for all participants combined; full and core set

	Extended set, all participants				*Core set, all participants*			
	M(β)	*SE*	*t*	*p*	*M*(β)	*SE*	*t*	*p*
(a) Broad-semantics model								
Intercept	0.00	0.03	0.00	.998	0.03	0.07	0.41	.682
Broad-range rule PO	**0.12**	**0.04**	**3.11**	**.002**	0.12	0.08	1.57	.116
Broad-range rule DO-1	**−0.09**	**0.04**	**−2.53**	**.011**	**−0.20**	**0.06**	**−3.23**	**.001**
Broad-range rule DO-2	**0.15**	**0.03**	**5.50**	**.000**	0.09	0.07	1.35	.177
Verb variance	0.18				0.12			
Participant variance	0.00				0.00			
(b) Narrow-semantics model								
Intercept	0.00	0.02	−0.01	.993	0.04	0.07	0.49	.625
Broad-range rule PO	**0.16**	**0.06**	**2.64**	**.008**	0.12	0.13	0.90	.370
Broad-range rule DO-1	−0.03	0.06	−0.59	.553	−0.10	0.10	−0.97	.331
Broad-range rule DO-2	−0.02	0.03	−0.55	.583	−0.01	0.09	−0.15	.877
SI speech	**0.17**	**0.04**	**4.23**	**.000**	0.06	0.08	0.73	.468
S2 mailing	**−0.16**	**0.05**	**−3.44**	**.001**	−0.15	0.08	−1.84	.065
S3 bequeathing	**−0.06**	**0.03**	**−1.93**	**.054**	−0.05	0.05	−1.00	.317
S4 motion	**0.15**	**0.03**	**4.79**	**.000**	0.08	0.07	1.27	.203
Verb variance	0.14				0.11			
Participant variance	0.00				0.00			
(c) Semantics + morphophonology								
Intercept	0.00	0.02	0.01	.994	0.06	0.08	0.72	.473
Broad-range rule PO	**0.21**	**0.06**	**3.41**	**.001**	0.14	0.14	1.00	.316
Broad-range rule DO-1	−0.07	0.06	−1.22	.221	−0.12	0.11	−1.08	.282
Broad-range rule DO-2	−0.01	0.03	−0.45	.656	−0.02	0.09	−0.22	.826
SI speech	**0.16**	**0.04**	**3.97**	**.000**	0.06	0.08	0.71	.478
S2 mailing	**−0.18**	**0.05**	**−4.00**	**.000**	−0.17	0.09	−1.93	.053
S3 bequeathing	**−0.08**	**0.03**	**−2.72**	**.007**	−0.06	0.06	−1.13	.258
S4 motion	**0.13**	**0.03**	**3.99**	**.000**	0.07	0.07	1.06	.289
Morphophonological constraint	**0.09**	**0.03**	**2.96**	**.003**	0.07	0.10	0.64	.520
Verb variance	0.14				0.11			
Participant variance	0.00				0.00			
(d) SEM + morphophonology + pre-emption								
Intercept	−0.10	0.02	−4.20	.001	−0.22	0.07	−3.04	.002
Broad-range rule PO	**0.16**	**0.06**	**2.84**	**.001**	0.04	0.10	0.37	.709
Broad-range rule DO-1	−0.04	0.05	−0.68	.476	0.04	0.08	0.51	.610
Broad-range rule DO-2	−0.03	0.03	−0.89	.338	0.00	0.06	−0.01	.993
SI speech	**0.15**	**0.04**	**4.21**	**.001**	0.03	0.06	0.47	.639
S2 mailing	**−0.19**	**0.04**	**−4.65**	**.001**	**−0.15**	**0.06**	**−2.52**	**.012**
S3 bequeathing	**−0.09**	**0.03**	**−3.37**	**.001**	**−0.09**	**0.04**	**−2.21**	**.027**
S4 motion	**0.08**	**0.03**	**2.64**	**.008**	0.00	0.05	0.08	.936
Morphophonological constraint	0.03	0.03	1.19	.172	−0.08	0.07	−1.13	.257
Pre-emption (BNC all texts)	**0.23**	**0.03**	**8.22**	**.001**	**0.28**	**0.04**	**6.36**	**.000**
Verb variance	0.11				0.05			
Participant variance	0.00				0.00			
(e) Semantics + morphophonology + pre-emption +entrenchment								
Intercept	−0.10	0.02	−4.21	.000	−0.25	0.07	−3.35	.001
Broad-range rule PO	**0.15**	**0.06**	**2.76**	**.006**	0.02	0.10	0.20	.838
Broad-range rule DO-1	−0.03	0.05	−0.62	.536	0.07	0.08	0.83	.407
Broad-range rule DO-2	−0.03	0.03	−0.94	.347	0.01	0.06	0.09	.926
SI speech	**0.15**	**0.04**	**4.21**	**.000**	0.01	0.06	0.13	.898
S2 mailing	**−0.19**	**0.04**	**−4.57**	**.000**	**−0.13**	**0.06**	**−2.21**	**.027**
S3 bequeathing	**−0.09**	**0.03**	**−3.32**	**.001**	−0.07	0.04	−1.81	.071

UNCORRECTED PROOF

Table 2 (*Continued*)

	Extended set, all participants					*Core set, all participants*			
	M(β)	*SE*	*t*	*p*		*M*(β)	*SE*	*t*	*p*
S4 motion	**0.08**	**0.03**	**2.63**	**.009**		0.01	0.05	0.15	.881
Morphophonological constraint	0.03	0.03	1.12	.261		−0.08	0.07	−1.10	.271
Pre-emption (BNC all texts)	**0.23**	**0.03**	**8.21**	**.000**		**0.30**	**0.04**	**6.60**	**.000**
Entrenchment (BNC all texts)	0.02	0.03	0.83	.407		0.08	0.06	1.41	.159
Verb variance	0.11					0.05			
Participant variance	0.00					0.00			
(f) Pre-emption-only model									
Intercept	−0.12	0.03	−4.58	.001		−0.25	0.06	−4.43	.001
Pre-emption (BNC all texts)	**0.28**	**0.03**	**9.88**	**.001**		**0.28**	**0.04**	**6.71**	**.001**
Verb variance	0.14					0.07			
Participant variance	0.00					0.00			
Model comparisons	AIC	logLik	χ^2	*p*	df	AIC	LogLik	χ^2	*P*
(a) Broad-semantics model (df = 7) vs	29811	−14898				7636	−3811		
(b) Narrow-semantics model (df = 11) vs	29753	−14865	65.71	.000	4	7638	−3808	6.53	.163
(c) Semantics + morphophonology (df = 12) vs	29746	−14861	8.85	.003	1	7640	−3808	0.28	.596
(d) Semantics + morphophonology + pre-emption (df = 13) vs	**29685**	**−14830**	**62.66**	**.000**	1	**7607**	**−3791**	**34.21**	**.000**
(e) Semantics + morphophonology + pre-emption + Ent (df = 14)	29687	−14829	0.68	.410	1	7607	−3790	2.25	.133
(d) Semantics + morphophonology + pre-emption (df = 13) vs	**29685**	**−14830**				**7607**	**−3791**		
(f) Pre-emption only	29753	−14871	83.5	.000	8	7612	−3801	20.24	.009
Age and interactions added to Model d (df = 33 for model, 20 for comparison)						**7532**	**−3733**	**116**	**.000**

predictor showed only a small (–0.09) negative correlation with the outcome variable, indicating a slight preference for DO- over PO-uses. Surprisingly, the *mailing* predictor, derived primarily from alternating (e.g., Transfer is mediated by a separation in time and space) and PO-only classes (B is something that C deserves, needs, or is worthy of), was also associated with a significant preference for DO- over PO-dative uses.

This pattern would seem to suggest that Pinker's (1989) proposal is accurate with regard to the types of fine-grained semantic features that determine verbs' dative argument-structure properties, but is inaccurate in its characterisation of precisely how they do so. One possibility is that the narrow-range classes proposed need to be redrawn. This would have no particular consequences for the theory; indeed Pinker (1989) acknowledges that the class assignments proposed are unlikely to be definitive. Another possibility, however, is that narrow-range classes are not quite the right way to characterise verbs' behaviour. This is an issue to which we return in the discussion.

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:21

In summary, when looking at all data combined (i.e., from 301 verbs for adults, 44 for children), optimal coverage is achieved by a model that contains broad- and narrow-range semantic predictors and pre-emption/entrenchment, but not the hypothesised morphophonological Latinate constraint.

Core set (44 verbs, all participants)

The observed finding of significant effects of verb semantics and pre-emption/entrenchment raises the question of whether all these predictors are important at all ages, or whether their relative importance differs with age. As a first step towards investigating this issue, we reran the analysis above, this time including only the core set of verbs ($N = 44$) rated by all age groups, and investigated whether adding age and its associated interaction terms to the model would yield significantly improved coverage (see rightmost columns of Table 2). Details of this analysis are presented in Section D of the online supplementary material. The important point for our purposes is simply that adding the interactions of age by each of the predictor variables significantly improved model coverage, with 11 significant interactions observed. Thus, there is ample evidence to suggest that the influence of these predictors changes with age. To investigate this pattern of development, we conducted the analysis described above for each age group separately.

By-age analyses (core set, 44 verbs)

Figures 2 and 3, respectively, illustrate (1) the mean preference for PO- over DO-dative sentences (difference scores) and (2) the mean rating for DO-dative sentences, by age group and verb, shown in the original units (points on the 5-point scale). Note that the core set includes both PO-only verbs and alternating verbs, but no DO-only verbs. The plots are arranged so that the acceptability of DO-dative uses, as rated by adults—(1) relative to PO-dative uses and (2) in absolute terms—increases travelling downwards. Verbs that are PO-only (non-alternating) according to Pinker (1989) and Levin (1993) are marked with an asterisk (*). All except three (*push*, *drop*, and *drag*) are in the top half of Figure 2, suggesting that these classifications are largely accurate. Visual inspection of Figure 2 suggests that, taking adult judgements as a reference, the 5- to 6-year-olds—but not the 9- to 10-year-olds—are making judgements that correspond to overgeneralisation errors. That is they display lower difference scores than adults for non-alternating verbs, but not for alternating verbs. A 2 × 3 (verb type × age) by-items ANOVA yielded a significant interaction [$F(2, 84) = 10.75$, $p < .001$, $\eta_p^2 = .20$], confirming this pattern. Planned comparisons revealed no by-age differences for the alternating verbs, but confirmed that, for the non-alternating verbs (using, for this analysis only, the Pinker/Levin classifications), mean difference scores were significantly lower for 5- to 6-year-olds ($M = 0.49$, $SE = 0.11$) than for both 9- to 10-year-olds ($M = 1.22$, $SE = 0.15$; $p < .001$) and adults ($M = 0.97$, $SE = 0.10$, $p = .001$). Unexpectedly, mean difference scores were slightly greater for 9- to 10-year-olds than for adults ($M = 0.97$, $SE = 0.10$, $p = .03$), indicating that these older children were—if anything—hypersensitive to overgeneralisation errors (or perhaps more willing to use the ends of the scale).

Importantly, the finding that the younger children's judgements did not differ from those of older children or adults for the alternating verbs allows us to be confident that these children were using the scale in a reasonably adult-like way, and, thus, that the lower difference scores for the non-alternating verbs reflect a genuine difference in the perceived ungrammaticality of errors. Indeed, visual inspection of the raw ratings (Figure 3) confirms that 5- to 6-year-olds were giving ratings that correspond to

UNCORRECTED PROOF

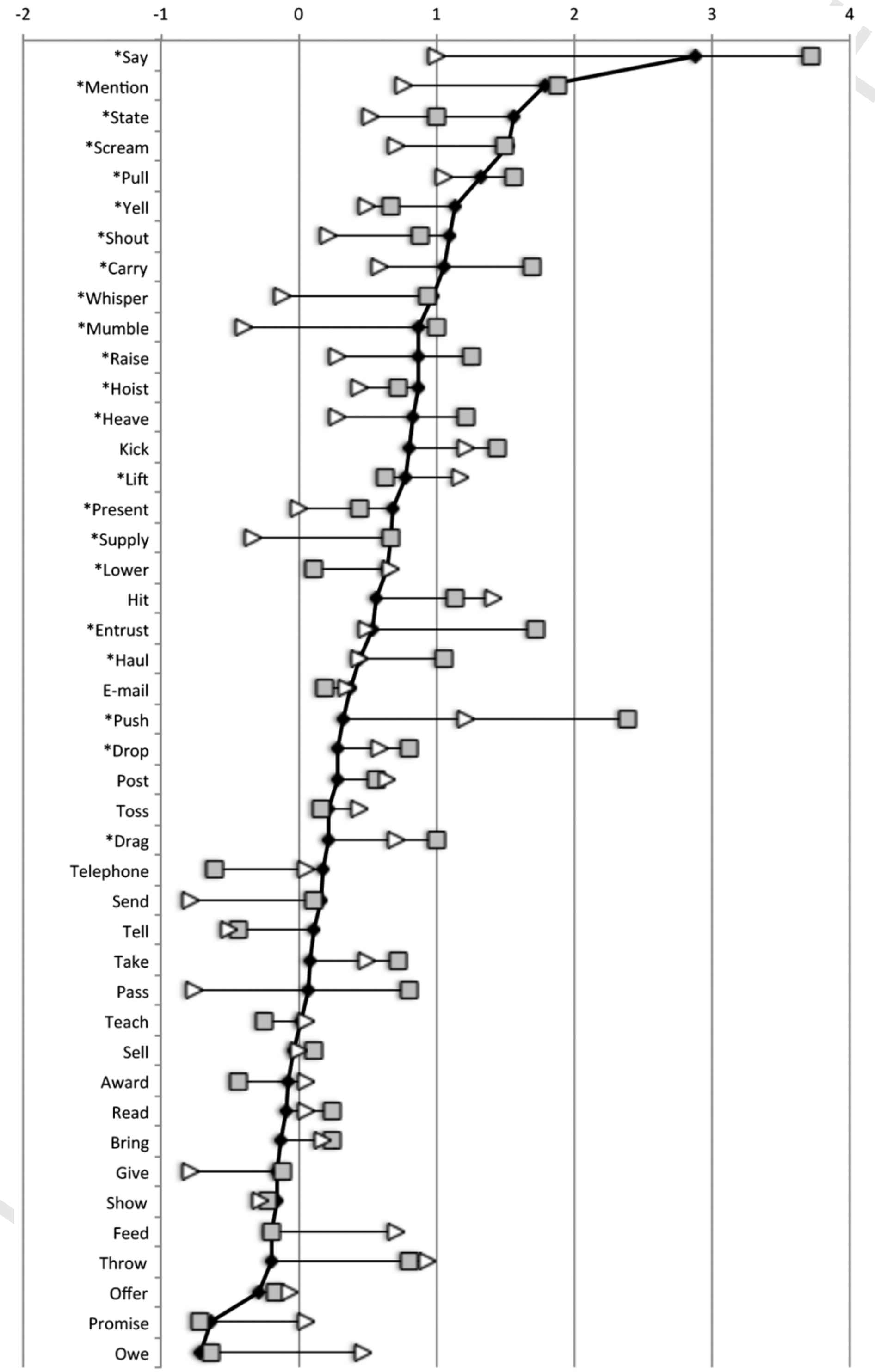

Figure 2. Mean by-verb and by-age difference scores (PO-dative minus DO-dative) in the original units (points on the 5-point scale), for the core set of 44 verbs (PO-only and alternating). Asterisks indicate PO-only verbs (i.e., DO-uses are ungrammatical and are predicted to receive high difference scores, representing a large dispreference for ungrammatical DO-dative vs grammatical PO-dative uses).

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:22

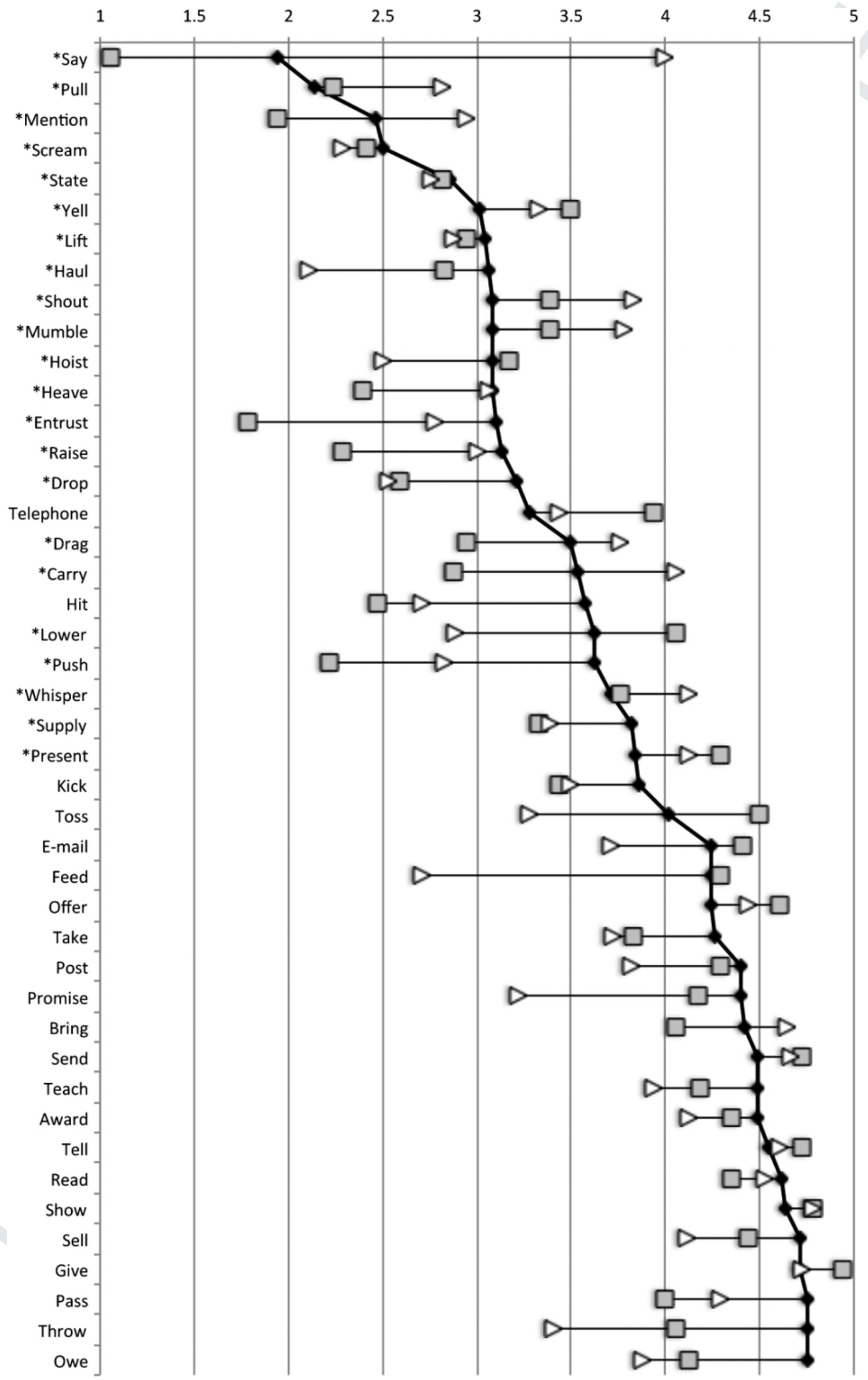

Figure 3. Mean by-verb and by-age raw ratings for DO-dative sentence in the original units (points on the 5-point scale), for the core set of 44 verbs (PO-only and alternating). Asterisks indicate PO-only verbs (i.e., DO-uses are ungrammatical and are predicted receive ratings towards the lower end of the scale).

overgeneralisation errors in production, for example, assigning ungrammatical DO-uses of *say*, *pull*, *mention*, *yell*, *shout*, and *mumble* ratings at least 0.5 points higher than those given by adults. A statistical analysis investigating the roles of the semantic and statistical predictors in these developmental changes is presented in a section that follows the by-age analyses presented here.

For these by-age analyses, the intention had been to remove any semantic predictor that loaded primarily on DO-only classes, which were not included in this core set. However, there was no semantic predictor of this type, presumably because only one of the original semantic features related to DO-only verbs (A makes some commitment that C can or cannot have B in the future). The morphophonology predictor was not included in this analysis, as native/Latinate status was not systematically manipulated in the core set (and a preliminary analysis found that it did not explain any variance for any age group). The by-age analyses are shown in Table 3.

Adults (core set only)

For adults (see Table 3, first column), the optimal model (c) included the narrow-range semantic predictors (of which two, *mailing* and *bequeathing*, were significant) and pre-emption/entrenchment. As for the main analysis, whichever of the two predictors is entered first explains significant variance, with the second adding no significant improvement ($\chi^2 = 0.68$, $p = .40$, when adding the residualised pre-emption predictor to an entrenchment-only model), though, in this case, entrenchment is closer than pre-emption to explaining significant additional variance when entered as a residualised predictor ($p = .06$ for the model comparisons). The broad-range semantic predictors were no longer significant once the narrow-range predictors were added to the model. However, even if the broad-range rules add nothing to the statistical model, this is not to say that they can be omitted from Pinker's (1989) theoretical account. Under this account, the broad-range rules specify the types of semantic features to which the narrow-range rules are sensitive. For example, the broad-range rule for the DO-dative specifies possession transfer, and the narrow-range classes different extents, kinds or manners of transfer. Thus, if we were to take a purely data-driven approach and omit the broad-range rules, we would be left with no explanation of why the learning mechanism that forms narrow-range classes is sensitive only to certain features. In the general discussion, we will outline an account that dispenses with the distinction between broad- and narrow-range semantics, and that hence renders moot the question of the relative contributions of each. Although numerous interactions were observed, they do not affect the observation that the broad- and narrow-range semantic classes, as well as entrenchment/pre-emption are important predictive factors for adults. Thus, we reserve discussion of these interactions (and the effect of adding by-participant slopes) for the online supplementary material (Section E and Table S1 of the online supplementary material).

Age 9–10 (core set only)

This analysis was exactly equivalent to the adult analysis described above. An alternative version of the analysis used entrenchment and pre-emption counts obtained from the post-Manchester corpus. Although it is debatable whether this corpus or the BNC is most representative of the speech heard by 9- to 10-year-olds, an identical pattern of results (and a very similar pattern of β, t, and p values) is obtained, whichever counts are used. We present the analysis based on the BNC counts, as this yields slightly better coverage of the data, presumably because the increased size of the

UNCORRECTED PROOF

AQ31

TABLE 3
Separate analysis for each age group

	Core set, adults				*Core set, age 9–10 years*				*Core set, age 5–6 years*			
	M(β)	*SE*	*t*	*p*	*M*(β)	*SE*	*t*	*p*	*M*(β)	*SE*	*t*	*p*
(a) Broad-semantics model												
Intercept	0.09	0.09	0.98	.328	−0.02	0.11	−0.22	.825	−0.03	0.05	−0.54	.592
Broad-range rule PO	0.02	0.09	0.21	.835	**0.23**	**0.11**	**2.07**	**.039**	**0.20**	**0.05**	**3.86**	**.000**
Broad-range rule DO-1	**−0.18**	**0.08**	**−2.37**	**.018**	**−0.24**	**0.09**	**−2.67**	**.008**	**−0.19**	**0.04**	**−4.67**	**.000**
Broad-range rule DO-2	0.06	0.09	0.67	.502	0.16	0.10	1.52	.129	**0.10**	**0.05**	**2.17**	**.030**
Verb variance	0.18				0.24				0.01			
Participant variance	0.00				0.00				0.00			
(b) Narrow-semantics model												
Intercept	0.10	0.09	1.16	.245	−0.01	0.11	−0.08	.939	−0.04	0.05	−0.84	.403
Broad-range rule PO	0.08	0.16	0.49	.626	0.22	0.20	1.11	.266	0.10	0.09	1.12	.262
Broad-range rule DO-1	−0.08	0.12	−0.63	.530	−0.06	0.15	−0.40	.692	**−0.18**	**0.07**	**−2.64**	**.008**
Broad-range rule DO-2	−0.11	0.10	−1.06	.291	0.02	0.13	0.12	.903	**0.13**	**0.06**	**2.31**	**.021**
SI speech	0.14	0.10	1.39	.164	0.04	0.12	0.31	.759	−0.06	0.05	−1.10	.270
S2 mailing	**−0.23**	**0.10**	**−2.39**	**.017**	**−0.26**	**0.12**	**−2.13**	**.033**	0.10	0.05	1.81	.071
S3 bequeathing	−0.09	0.06	−1.35	.177	−0.07	0.08	−0.89	.376	0.02	0.03	0.49	.623
S4 motion	0.11	0.08	1.48	.139	0.08	0.10	0.86	.392	0.03	0.04	0.70	.485
Verb variance	0.15				0.23				0.00			
Participant variance	0.00				0.00				0.00			
(c) Semantics + pre-emption												
Intercept	−0.14	0.07	−1.95	.052	−0.30	0.09	−3.22	.001	−0.11	0.06	−1.88	.080
Broad-range rule PO	0.01	0.11	0.09	.928	0.14	0.14	0.98	.329	0.08	0.09	0.92	.380
Broad-range rule DO-1	0.06	0.09	0.64	.520	0.10	0.12	0.87	.382	**−0.14**	**0.07**	**−2.06**	**.048**
Broad-range rule DO-2	−0.10	0.07	−1.41	.158	0.02	0.09	0.24	.810	**0.13**	**0.06**	**2.33**	**.026**
SI speech	0.10	0.07	1.42	.157	−0.01	0.09	−0.07	.941	−0.07	0.05	−1.27	.216
S2 mailing	**−0.25**	**0.07**	**−3.63**	**.000**	**−0.28**	**0.09**	**−3.18**	**.002**	0.09	0.05	1.71	.104
S3 bequeathing	**−0.13**	**0.05**	**−2.96**	**.003**	**−0.13**	**0.06**	**−2.20**	**.028**	0.00	0.04	0.12	.836
S4 motion	0.01	0.06	0.22	.829	−0.04	0.07	−0.57	.572	0.00	0.04	0.03	.958
Pre-emption (BNC all texts)	**0.31**	**0.05**	**6.08**	**.000**	**0.37**	**0.07**	**5.73**	**.000**	**0.09**	**0.04**	**2.16**	**.042**
Verb variance	0.07				0.10				0.00			
Participant variance	0.00				0.00				0.00			

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:23

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:23

UNCORRECTED PROOF

Table 3 (*Continued*)

	Core set, adults					*Core set, age 9–10 years*				*Core set, age 5–6 years*			
	M(β)	*SE*	*t*	*p*		*M*(β)	*SE*	*t*	*p*	*M*(β)	*SE*	*t*	*p*
(d) Semantics + pre-emption + Entrenchment													
Intercept	−0.19	0.08	−2.49	.013		−0.31	0.10	−3.02	.003	−0.14	0.06	−2.28	.023
Broad-range rule PO	−0.02	0.11	−0.15	.884		0.14	0.15	0.94	.348	0.06	0.09	0.72	.473
Broad-range rule DO-1	0.10	0.09	1.09	.278		0.10	0.12	0.86	.389	−0.12	0.07	−1.62	.106
Broad-range rule DO-2	−0.09	0.07	−1.31	.189		0.02	0.10	0.24	.808	**0.14**	**0.06**	**2.43**	**.015**
SI speech	0.07	0.07	1.00	.320		−0.01	0.09	−0.10	.923	−0.09	0.06	−1.59	.112
S2 mailing	**−0.22**	**0.07**	**−3.25**	**.001**		**−0.28**	**0.09**	**−3.03**	**.003**	**0.11**	**0.06**	**2.00**	**.046**
S3 bequeathing	**−0.11**	**0.05**	**−2.46**	**.014**		**−0.13**	**0.06**	**−2.05**	**.040**	0.02	0.04	0.53	.593
S4 motion	0.02	0.06	0.32	.749		−0.04	0.07	−0.55	.582	0.01	0.04	0.12	.908
Pre-emption	**0.33**	**0.05**	**6.51**	**.000**		**0.37**	**0.07**	**5.48**	**.000**	**0.10**	**0.04**	**2.48**	**.013**
Entrenchment	0.12	0.07	1.77	.077		0.01	0.09	0.11	.912	0.08	0.05	1.52	.129
Verb variance	0.06					0.11				0.00			
Participant variance	0.00					0.00				0.00			
(e) Pre-emption only													
Intercept	−0.25	0.06		.001		−0.33	0.08	−3.93	.000	−0.13	0.05	−2.46	.014
Pre-emption	**0.28**	**0.04**	**6.71**	**.001**		**0.37**	**0.06**	**5.96**	.000	**0.15**	**0.04**	**3.74**	**.000**
Verb variance	0.07					0.14				0.02			
Participant variance	0.00					0.00				0.00			
Model comparisons	AIC	logLik	χ^2	*p*	df	AIC	logLik	χ^2	*p*	AIC	logLik	χ^2	*p*
(a) Broad-semantics model (df = 7) vs	3522	−1754				1924	−955			2135	−1060		
(b) Narrow-semantics model (df = 11) vs	3519	−1749	11.10	.025	4	1926	−952	6.46	.168	2139	−1059	3.40	.494
(c) Semantics + pre-emption (df = 12) vs	**3490**	**−1733**	**31.47**	**.000**	**1**	**1899**	**−938**	**28.78**	**.000**	**2136**	**−1056**	**5.37**	**.020**
(d) Semantics + pre-emption + Entrenchment (df = 13)	3488	−1731	3.62	.057	1	1901	−938	0.00	1.000	2136	−1055	2.34	.126
(c) Semantics + pre-emption (df = 12) vs	**3490**	**−1733**				**1899**	**−938**			**2136**	**−1056**		
(e) Pre-emption only (df = 5)	3502	−1746	26.22	.000	7	1902	−946	16.92	.018	2140	−1065	18.13	.011

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:23

corpus more than compensates for any lack of representativeness. The same is true for the analysis of the younger children's data presented in the subsequent section.

The results of this analysis are shown in the middle columns of Table 3. The results were almost identical to those of the adults, with significant effects observed for two narrow-range predictors (*mailing* and *bequeathing*) and pre-emption (indeed, the β and *t* values are very similar across the two age groups). The only major difference is that, whilst adding the residualised entrenchment predictor to a model including pre-emption explained precisely no additional variance ($\chi^2 = 0.00$, $p = 1$, see Table 3), adding the residualised pre-emption predictor to a model including entrenchment yielded a significant improvement ($\chi^2 = 5.30$, $p = .02$). Thus, it may be that for the older children, pre-emption is slightly more important than entrenchment, with the reverse true for adults. However, given the fact that the two measures are so highly correlated ($r = 0.90$, $p < .001$), it is probably sensible to conclude no more than that both predictors are important for both age groups. Finally, it was again verified that removing the semantic predictors for a pre-emption-only model yielded a significantly worse fit (see final two rows of Table 3). The effect of adding interactions and random slopes is explored in Section E and Table S1 of the online supplementary material (again, these do not affect our overall conclusions).

Age 5–6 (core set only)

The younger children (see Table 3, right hand columns) displayed a very different pattern to that observed for the two older groups. For these children, two of the three broad-range semantic predictors, but none of the four narrow-range semantic predictors, were associated with significant effects. This is precisely the pattern predicted by an account under which children first learn broad-range semantic constraints on the constructions and hone these constraints until relatively late in childhood (presumably somewhere between 6 and 9 years), such as that of Pinker (1989). There was nothing to choose between pre-emption and entrenchment; whichever predictor was entered first explained a significant proportion of variance, with the other explaining no significant additional variance ($\chi^2 = 0.29$, $p = .59$ for the comparison between the entrenchment models with and without residualised pre-emption). Finally, it was again demonstrated that removing the semantic predictors from the model yielded a significantly worse fit to the data (see final two rows of Table 3). The effect of adding interactions and random slopes is explored in Section E and Table S1 of the online supplementary material (again, these do not affect our overall conclusions).

Developmental analysis

The analysis above investigates the roles of the broad- and narrow-range semantic predictors, entrenchment and pre-emption at three snapshots of development, but does not directly investigate their role in development. For the final analysis, we, therefore, calculated difference-change scores by calculating the mean by-verb difference score for each age group and subtracting the scores given by (1) 5- to 6-year-olds from 9- to 10-year-olds, (2) 9- to 10-year-olds from adults, and (3) 5- to 6-year-olds from adults. We then ran by-items regression analyses to investigate how the broad- and narrow-range semantic predictors, entrenchment, and pre-emption predict these *developmental changes* in participants' ratings (as measured by the difference-change scores). The results of these analyses are shown in Table 4. Note that, because

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:24

AQ32

TABLE 4
Analysis of developmental-change scores

	Change scores 5–6 to 9–5-year-olds				*Change scores 9- to 10-year-olds to adults*				*Change scores 5- to 6-year-olds to adults*			
	M	*SE*	*t*	*p*	*M*	*SE*	*t*	*p*	*M*	*SE*	*t*	*p*
(a) Broad-semantics model												
Intercept	0.33	0.16	2.08	.044	−0.01	0.10	−0.09	.932	0.32	0.14	2.30	.027
Broad-range rule PO	0.05	0.16	0.29	.770	**−0.35**	**0.10**	**−3.46**	**.001**	**−0.30**	**0.14**	**−2.13**	**.039**
Broad-range rule DO-1	−0.06	0.13	−0.42	.677	0.13	0.08	1.58	.122	0.08	0.12	0.65	.518
Broad-range rule DO-2	0.11	0.15	0.73	.472	−0.17	0.09	−1.86	.070	−0.07	0.13	−0.51	.615
(b) Narrow-semantics model												
(Intercept)	0.38	0.14	2.74	.009	−0.02	0.10	−0.19	.853	**0.36**	**0.12**	**3.15**	**.003**
Broad-range rule PO	0.24	0.25	0.96	.342	−0.26	0.18	−1.42	.165	−0.01	0.21	−0.07	.946
Broad-range rule DO-1	0.20	0.20	1.00	.325	0.00	0.14	0.04	.972	0.20	0.16	1.24	.224
Broad-range rule DO-2	−0.18	0.17	−1.11	.275	−0.20	0.12	−1.67	.103	**−0.38**	**0.14**	**−2.80**	**.008**
SI speech	0.19	0.16	1.24	.224	0.14	0.11	1.21	.233	**0.33**	**0.13**	**2.55**	**.015**
S2 mailing	**−0.61**	**0.16**	**−3.93**	**.000**	0.07	0.11	0.61	.548	**−0.55**	**0.13**	**−4.22**	**.000**
S3 bequeathing	−0.14	0.10	−1.37	.178	−0.02	0.07	−0.26	.796	−0.16	0.08	−1.89	.068
S4 motion	0.07	0.12	0.54	.595	0.04	0.09	0.50	.618	0.11	0.10	1.09	.285
(c) Semantics + pre-emption												
Intercept	0.08	0.14	0.54	.591	0.07	0.12	0.63	.531	0.15	0.12	1.22	.230
Broad-range rule PO	0.16	0.21	0.75	.460	−0.23	0.18	−1.29	.206	−0.07	0.19	−0.39	.702
Broad-range rule DO-1	**0.37**	**0.17**	**2.14**	**.039**	−0.05	0.14	−0.32	.748	**0.32**	**0.15**	**2.12**	**.041**
Broad-range rule DO-2	−0.18	0.14	−1.27	.213	−0.20	0.12	−1.72	.095	−0.38	0.12	−3.08	.004
SI speech	0.15	0.13	1.11	.276	0.15	0.11	1.36	.184	0.30	0.12	2.56	.015
S2 mailing	**−0.64**	**0.13**	**−4.84**	**.000**	0.07	0.11	0.68	.504	**−0.56**	**0.12**	**−4.85**	**.000**
S3 bequeathing	**−0.20**	**0.09**	**−2.32**	**.026**	0.00	0.07	0.00	.997	**−0.20**	**0.08**	**−2.63**	**.013**
S4 motion	−0.06	0.11	−0.57	.575	0.08	0.09	0.91	.367	0.02	0.10	0.23	.818
Pre-emption (BNC all texts)	**0.39**	**0.10**	**4.01**	**.000**	−0.12	0.08	−1.45	.157	**0.27**	**0.08**	**3.16**	**.003**

UNCORRECTED PROOF

UNCORRECTED PROOF

Table 4 (*Continued*)

	Change scores 5–6 to 9–5-year-olds				*Change scores 9- to 10-year-olds to adults*				*Change scores 5- to 6-year-olds to adults*			
	M	*SE*	*t*	*p*	*M*	*SE*	*t*	*p*	*M*	*SE*	*t*	*p*
(d) Semantics +pre-emption + Entrenchment												
(Intercept)	0.11	0.15	0.78	.441	0.05	0.14	0.40	.69	0.14	0.13	1.09	.284
Broad-range rule PO	0.19	0.22	0.86	.396	−0.24	0.18	−1.28	.21	−0.08	0.19	−0.41	.687
Broad-range rule DO-1	0.33	0.18	1.84	.074	−0.03	0.15	−0.22	.82	**0.33**	**0.16**	**2.08**	**.045**
Broad-range rule DO-2	−0.19	0.14	−1.33	.192	−0.2	0.12	−1.70	.10	**−0.38**	**0.12**	**−3.02**	**.005**
SI speech	0.18	0.14	1.29	.207	0.15	0.12	1.33	.19	**0.29**	**0.12**	**2.40**	**.022**
S2 mailing	**−0.66**	**0.14**	**−4.91**	**.000**	0.06	0.11	0.56	.58	**−0.56**	**0.12**	**−4.61**	**.000**
S3 bequeathing	**−0.22**	**0.09**	**−2.47**	**.019**	−0.02	0.07	−0.22	.83	**−0.20**	**0.08**	**−2.44**	**.020**
S4 motion	−0.07	0.11	−0.61	.544	0.06	0.09	0.69	.50	0.02	0.10	0.24	.811
Pre-emption	**0.38**	**0.10**	**3.91**	**.000**	−0.04	0.05	−0.80	.43	**0.27**	**0.09**	**3.12**	**.004**
Entrenchment	−0.12	0.13	−0.90	.377					0.03	0.12	0.22	.831
(e) Pre-emption only												
Intercept	**0.07**	**0.14**	**0.48**	**.635**	−0.06	0.10	−0.58	.569	0.01	0.14	0.05	.963
Pre-emption	**0.30**	**0.10**	**2.95**	**.005**	−0.14	0.08	−1.79	.081	0.16	0.10	1.62	.114
Model comparisons												
	df	*SS*	*F*	*p*	df	*SS*	*F*	*p*	df	*SS*	*F*	*p*
(a) Broad-semantics model vs												
(b) Narrow-semantics model vs	**4.00**	**7.62**	**6.03**	**.001**	4.00	1.07	1.22	.321	**4.00**	**7.39**	**7.37**	**.000**
(c) Semantics + pre-emption vs	**1.00**	**5.03**	**15.95**	**.000**	1.00	0.47	2.14	.153	**1.00**	**2.43**	**9.70**	**.004**
(d) Semantics + pre-emption + Ent	1.00	0.25	0.80	.377	1.00	0.37	1.70	.201	**1.00**	**0.01**	**0.05**	**.831**
(c) Semantics + pre-emption vs												
(e) Pre-emption only (df = 5)	**−7.00**	**−8.97**	**4.08**	**.002**	−7.00	−3.47	2.22	.056	−7.00	−11.15	6.53	.000
(c) Semantics + pre-emption vs												
(c) Semantics + pre-emption + interactions	19.00	7.06	1.52	.203	19.00	4.83	1.36	.271	19.00	5.96	1.95	.091

we are interested in how children's judgements change with age, these analyses use raw, untransformed ratings (as opposed to *z*-scores in the main analysis).

The first immediately apparent finding is that none of the semantic or statistical factors are significant predictors of any change between the older children and adults (with the sole exception of the broad-range rule for the PO-dative in the model that contains only broad-range rules). This is probably a reflection of the fact that there do not seem to be any important developmental changes between age 9–10 and adulthood; a finding reflected in the ANOVA reported at the beginning of the previous section.

Accordingly the models that investigate the changes between (1) age 5–6 and 9–10 and (2) age 5–6 and adulthood are virtually identical. In each case, the optimal model includes the broad- and narrow-range semantic predictors and pre-emption, but not entrenchment (though as for the previous analyses, entrenchment is always a significant predictor if entered before pre-emption; though these models are not shown). Neither does the optimal model contain any interactions. Turning to the individual predictors, the broad-range rule for the DO-dative and the pre-emption measure are positive predictors of the change scores. For the narrow-range predictors *mailing* and *bequeathing*, the sign of the difference is negative, but this is only because these narrow-range classes seem to favour DO- over PO-uses (and were also negative in all previous analyses). Thus, across verbs, the greater the compatibility of the verb with (1) the broad-range rule for the DO-dative, the narrow-range predictors of (2) mailing and (3) bequeathing, and (4) the greater the availability of pre-empting alternatives, the greater the magnitude of change between younger and older participants' difference scores. This is perhaps the most direct evidence yet of roles for broad- and narrow-range semantics and pre-emption, as we are aware of no previous studies demonstrating that these factors play a role in developmental *change*, as opposed simply to predicting by-verb variance at a particular age. It is important to emphasise, however, that the semantic ratings were collected from adults only. Thus, the significant semantic effects here indicate that the extent to which sentence-acceptability judgements are predicted by adult semantic ratings increases with age. We would suggest that the most likely explanation for this finding is that, as development proceeds, children's semantic verb representations are nearing the adult state (which is why semantics predicts judgements equally well for 9–10-year-olds and adults). However, we acknowledge that in order to test this prediction more directly, it would be necessary to obtain semantic ratings (or some other measure of semantic knowledge, e.g., using a comprehension task; e.g., Gropen, Pinker, Hollander, & Goldberg, 1991) from children.

Summary of results

The overall pattern observed is summarised below and in Table 5.

- Both **pre-emption** and **entrenchment** are significant predictors at every age, and of the developmental changes observed between ages 5–6 and 9–10/adulthood. The two predictors largely explain overlapping variance, and hence there is little to choose between them. There is some evidence to suggest that pre-emption explains additional variance beyond entrenchment for the 9- to 10-year-olds, and for the change between ages 5–6 and 9–10, with the reverse (marginally) true for adults. However, since the two predictors were so highly correlated, it would seem wise to await further investigation (most likely manipulating the two factors experimentally) before drawing any firmer conclusions regarding the relative contributions of the two mechanisms.

UNCORRECTED PROOF

AQ35

TABLE 5
Summary of developmental pattern

		Core set			
Main (cross-sectional) analysis	*Extended set*	*All ages*	*Adults*	*Age 9–10*	*Age 5–6*
Broad-range rules (at least one of three)	Yes (1)	Yes[a]	Yes[a]	Yes[a]	Yes (2)
Narrow-range classes (at least one of four)	Yes (4)	Yes (2)	Yes (2)	Yes (2)	No
Morphophonology	Yes[a]	No		n.a.	
Pre-emption	Yes	Yes[b]	Yes[b]	Yes	Yes[b]
Entrenchment	Yes[b]	Yes[b]	Yes[c]	Yes[b]	Yes[b]
Developmental analysis		5–6 to 9–5	9–5 to adult	5–6 to adult	
Broad-range rules (at least 1 of 3)		Yes (1)	Yes[a]	Yes (1)	
Narrow-range classes (at least 1 of 4)		Yes (2)	No	Yes (2)	
Pre-emption		Yes	No	Yes[b]	
Entrenchment		Yes[b]	No	Yes[b]	

[a]Significant predictor independently if entered before – but not after – narrow range semantic predictors.
[b]Significant predictor if entered before the other statistical predictor, but does not explain additional variance if entered after.
[c]Significant predictor if entered before pre-emption, and explains a marginally non-significant proportion of additional variance if entered after ($p = .06$).

- **Narrow-range semantic verb classes** have an effect for the two older groups, but not the 5–6-year-olds. Consequently, they are a significant predictor of developmental change between ages 5–6 and 9–10/adulthood.
- The effect of the **broad-range semantic rules** is less easy to discern because these serve to motivate the narrow-range classes, even when they are not significant predictors. Essentially, the pattern seems to be that semantics exerts its effects at a broad-brush level initially, but in an increasingly fine-grained manner as development proceeds. Consequently, broad-range semantics is also a predictor of developmental change.
- The **Latinate morphophonological constraint** (manipulated only for the full set of verbs, which was not rated by children) did not explain additional variance above and beyond semantics and frequency, but was a significant predictor of judgements when assessed independently. Taken together with the finding of AQ12 studies using novel pseudo-Latinate verbs (e.g., Ambridge et al., 2012; Gropen, Pinker, Hollander, Goldberg, Wilson, 1989), this finding suggests that this constraint is psychologically real for adults, but probably not for children as old as 9–10 years.

DISCUSSION

In the present study, adults and children rated the acceptability of 301 and 44 verbs, respectively, in PO- and DO-dative constructions, in order to test competing theoretical accounts of how children avoid—or retreat from—dative overgeneralisation errors (e.g., **I said her ''no''*). A significant effect of pre-emption, whereby the use of a verb in the PO-dative construction constitutes evidence that DO-dative uses are not permitted (and vice-versa) was observed for all ages, and also when looking at developmental change. A significant effect of entrenchment, whereby the use of a verb

UNCORRECTED PROOF

in any construction constitutes evidence that nonattested PO- or DO-dative uses are not permitted was also observed for all ages (and, again accounted for developmental change). Although pre-emption and entrenchment largely explained the same variance, a significant additional effect of pre-emption was observed (1) for the older children, (2) when looking across the extended set of 301 verbs (as opposed to the core set of 44 verbs), and (3) when investigating developmental change between ages 5–6 and 9–10. For adults, the reverse was (almost) true, with entrenchment explaining a marginally significant proportion of variance beyond that accounted for by pre-emption. Verbs' semantic properties—as rated by independent adult raters—explained additional variance at all ages, with the relative influence of broad- and narrow-range semantic properties decreasing and increasing respectively with age. Finally, there was some support for a proposed morphophonological constraint that prohibits Latinate verbs from appearing in the DO-dative (e.g., **I suggested her the trip*), but this had no effect above and beyond semantics and pre-emption, suggesting that it may be learned on a verb-by-verb basis.

In the introduction, we outlined four possibilities for the relationship between semantics-based and statistics-based accounts: *Semantics only* (Ambridge et al., 2009; Pinker, 1989), *Statistics only* (Stefanowitsch, 2008), *Semantics with statistics for exceptions* (Bowerman, 1988; Boyd & Goldberg, 2011; Braine & Brooks, 1995), and *Statistics then semantics* (Perfors et al., 2010, 2011; Tomasello, 2003; Wonnacott, 2011). It is clear that neither semantics-only nor statistics-only accounts can explain the present dataset. Though different predictors varied in their relative importance with age, both semantics- and statistics-based predictors accounted for unique variance for every age group, by even the most stringent statistical criterion (i.e., the AIC measure, which strictly penalises the model for the introduction of each additional predictor). Perhaps more importantly, both semantics- and statistics-based predictors were needed in order to account for developmental change.

Neither do the findings provide support for a statistics-then-semantics account. Of course, the youngest children studied were—at 5–6 years—relatively old, and these findings do not rule out the possibility that children may display statistics-based but not semantics-based effects before this age (as proposed explicitly by Tomasello, 2003). However, as previous studies that have demonstrated pre-emption/entrenchment before semantics have used familiar verbs to assess the former and novel verbs to assess the latter (see Introduction section), we would argue that whether or not statistics exerts its effects before semantics remains an open question.

An account proposing semantics with statistics for exceptions (Boyd & Goldberg, 2011), would—at first glance—appear to be well supported by the present findings. Even with all semantic and morphophonological predictors in the model, pre-emption/entrenchment explained an additional proportion of unique variance.

In fact, however, detailed comparison of the observed ratings and the ratings predicted by the combined semantic and morphophonological predictors suggest that there are very few truly arbitrary exceptions. Table S2 in the supplementary online material presents the residuals from the model for all verbs that includes the semantic and morphophonological predictors (i.e., model d from the left hand column of Table 2). These represent, for each verb, the extent to which participants' preference for PO- over DO-dative uses is larger (left hand column) or smaller (right hand column, indicating a preference for DO- over PO-dative uses) than would be predicted on the basis of the semantic and morphophonological predictors, in standard deviation units. The usual criterion for an outlier (i.e., here, a verb whose use in the PO-/DO-dative is less acceptable than would be predicted from its semantics and

UNCORRECTED PROOF

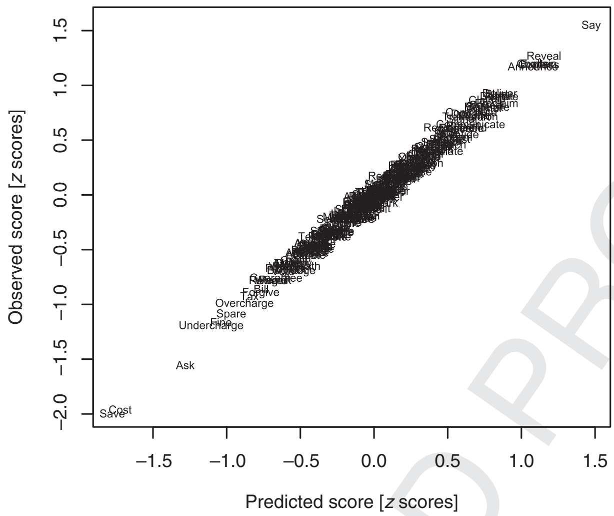

Figure 4. Mean observed ratings vs ratings predicted by a model including semantic and morphophonological predictors.

morphophonology) is an item with a residual of greater than 2.0 SDs. Table S2 shows that no verb has a residual of larger than 0.25 SDs. For any given verb, the degree of preference for a PO- over a DO-usage (or vice-versa) can be predicted almost perfectly by its semantic and morphophonological properties. Indeed, if the observed ratings are plotted against the predicted ratings (see Figure 4), most individual verbs are impossible to make out, because the points lie within such a narrow band that each is overplotted by many others.[6]

If there are no true exceptions, why did the present study find effects of pre-emption/entrenchment, even after controlling for semantics and morphology? We suggest that pre-emption and entrenchment exert their effects not as mechanisms for learning arbitrary exceptions, but by modulating semantics-based effects (hence the numerous interactions observed between the two types of predictor; see Table S1 in the online supplementary material). Below, we outline a more concrete proposal for precisely how pre-emption, entrenchment, and semantics can work together in this way. This is not to deny that pre-emption and/or entrenchment could, in principle, serve to rule out any exceptions to semantic and morphophonological constraints, but on the evidence of the present study, there is no good evidence to suggest that such exceptions exist, at least in the domain of restrictions on the dative.

Overall there was little to choose between pre-emption and entrenchment. The two predictors were highly correlated and, in most cases, appeared to explain similar variance, with each predictor explaining little or no additional variance beyond that explained by the other. If anything, in the present study, pre-emption appeared to be a slightly more important predictor than entrenchment; a pattern opposite to that AQ13 observed in Ambridge et al.'s (2012) analogous study involving the locative constructions. One possibility is that any discrepancy is due simply to methodological differences. For example, the previous study used a much smaller corpus (ICE-GB),

[6]Note that, to our knowledge, no researcher has proposed a *statistics, with semantics/morphophonology for exceptions* model. Such an account would claim that there exist verbs that appear in (for example) the PO-dative construction only (statistics), but whose semantic/morphophonological properties are such that speakers would also consider them to be grammatical in the DO-dative construction.

UNCORRECTED PROOF

meaning that both the pre-emption and entrenchment counts were less reliable. A more interesting possibility is that the balance between pre-emption and entrenchment differs according to the particular generalisation in question, and may depend on the relative frequency of the verb in potentially entrenching vs potentially pre-empting constructions.

The possibility that we favour, however, is that there is no clear divide between entrenchment and pre-emption: both are statistical-learning mechanisms that are sensitive to frequency of a particular verb in particular constructions; all that varies is the relative weighting of the different constructions. This would also explain why the two predictors account for largely the same variance.

Below, we outline in more detail an account that incorporates the type of statistical learning mechanism that we have in mind, and also yields effects of semantics and morphology. First, it is important to consider whether it would instead be possible to modify Pinker's (1989) semantic verb class account to accommodate statistical-learning effects. For example, although Pinker (1989) does not discuss frequency effects, it would be possible to add the assumption that lower frequency verbs take longer to be assigned to the relevant semantic class, and hence are more susceptible to error. However, there are two problems with such a proposal. First, whilst it could explain general frequency effects (i.e., entrenchment), this account would not explain pre-emption. Second, the present study demonstrates that even adults, whose semantic classes are presumably fully formed, display a probabilistic pattern of acceptability judgements. This does not sit comfortably with an account under which verbs are definitively assigned to PO-only, DO-only, or alternating classes. Furthermore, the semantic properties identified as predictive of participants' judgements do not appear to neatly classify verbs into these three categories, but rather cut across apparent PO-only, DO-only, and alternating verbs (see Table 1).

What is needed, then, is a hybrid account that yields effects of verb semantics, in a probabilistic as opposed to class-based manner, as well as effects of pre-emption and entrenchment. Below, we briefly outline one attempt at such an account, based on a proposal outlined in more detail elsewhere (Ambridge & Lieven, 2011; Ambridge, Pine & Rowland, 2012; Ambridge, Pine, Rowland, & Chang, 2012; Ambridge et al., 2011).

We assume that speakers acquire a construction by abstracting across concrete exemplars of the construction that they have stored in memory (an assumption shared by most usage- and exemplar-based accounts of acquisition; e.g., Bod, 2009; Gahl & Yu, 2006; Goldberg, 1995; Langacker, 1987; Tomasello, 2003). In this way, learners of English will acquire two dative constructions:

PO-dative: [A ("GIVER")] [B("ACTION")] [C ("THEME")] to [D ("RECIPIENT")]
(e.g., [John] [sent] [the package] to [Sue]

DO-dative: [P ("GIVER")] [Q("ACTION")] [R("RECIPIENT")] [S ("THEME")]
(e.g., [John] [sent] [Sue] [the package])

The slots are labelled with letters rather than semantic or syntactic categories in order to reflect the assumption that each slot in each construction exhibits its own probabilistic constellation of semantic (and morphophonological, etc.) properties (though traditional semantic labels are also given for ease of reference). For example, the [C] and [S] ("THEME") slots have different—though somewhat overlapping—properties. The properties of a particular slot are a weighted average of the properties shared by the items that have appeared in this position in the input

UNCORRECTED PROOF

utterances that gave rise to the construction, and may be semantic, phonological, or discourse-pragmatic in nature (there may be other possibilities too). For example, the [Q] ("ACTION") slot in the DO-dative construction probabilistically exhibits the relevant semantic properties highlighted in the present study (e.g., possession transfer), as well as the morphophonological property of being monosyllabic/Germanic. The [R] slot has the semantic property of being a potential possessor of S and discourse-pragmatic properties such as being light, given and pronominal, as opposed to heavy, new and realised as a full NP (e.g., Bresnan et al., 2007).

The key assumption of the present account is that overgeneralisation errors of the type investigated in the present study reflect the use of an item (here, a verb) in a slot with which it is less than optimally compatible in terms of its semantic, morphophonological or other properties. Although this assumption is shared with many previous proposals (e.g., Bowerman, 1981; Braine & Brooks, 1995; Kay & Fillmore, 1999; Langacker, 2000), the present account goes further in assuming (like the account of Pinker, 1989) that all argument-structure overgeneralisation errors can be understood in this way; that is, there are no arbitrary exceptions that must be learned on a purely statistical basis.

Under this account, children produce errors because they have yet to fully acquire **AQ14** the semantic properties of particular verbs and/or particular construction slots,[7] and so insert verbs into slots with which they are less than fully compatible for adult speakers (e.g., *I said her "no"*). Errors cease as children refine their knowledge of the properties of particular verbs and construction slots, and so no longer insert verbs into slots with which they are less than optimally compatible.[8] This would explain why, in the present study, even young children's judgements are sensitive to broad-range semantic properties, with sensitivity to more narrow-range semantic properties increasing with development.

In order to also accommodate pre-emption and entrenchment, the present account adopts from, amongst others, Bates and MacWhinney (1987), the notion of competition; more specifically, competition between constructions for the right to express the speaker's desired message (e.g., Bowerman, 1981). Suppose, for example, that the speaker wishes to express the meaning that "Marge caused Homer to have a box by pulling a box towards him", and has decided to use the NPs *Marge*, *the box*, and *Homer*, and the verb *pulled*. All constructions in the speakers' inventory will compete for activation (i.e., for the right to express the message). Each construction has a resting activation level proportional to its overall frequency. This activation level is boosted depending on the extent to which each construction exhibits (1) semantic/morphophonological fit with the items in the message (as described above), (2) relevance, and (3) item-in-construction frequency.

[7] One study that predicted, but did not find, an effect of construction-slot semantics is that of Matthews and Bannard (2010). These authors found that two- and three-year old children showed no increased ability to repeat a familiar string when the relevant slot-filler was semantically similar to the corresponding filler in a high frequency (and hence presumably well-learned) string. For example, considering the familiar string *a piece of toast,* children made just as many errors when repeating *a piece of meat* as *a piece of brick*. However, one important difference from the present study is that the semantic properties of the fillers (e.g., *brick*), though *atypical* of the semantic properties of the slot, were not *incompatible* with them. Another is that the present participants were considerably older; and it would seem likely that 2–3-year-olds have learned less about construction semantics.

[8] See Lidz, Gleitman, and Gleitman (2004), Naigles, Fowler, and Helm (1992), and Naigles, Gleitman, and Gleitman (1993) for studies investigating children's interpretations of utterances with mismatching verb and construction semantics.

UNCORRECTED PROOF

Relevance is the extent to which each of the competing constructions expresses the desired message. For the present example ("Marge caused Homer to have a box by pulling a box towards him"), both the PO-dative (which would yield *Marge pulled the box to Homer*) and the DO-dative (**Marge pulled Homer the box*) score high on relevance, as they express all components of the intended message. Which of the two constructions is more relevant will depend on whether the speaker wishes to emphasise the movement of the box from one location to another (PO-dative) or that Homer came to possess the box (DO-dative). The transitive (which would yield *Marge pulled the box*) and intransitive (**The box pulled*) score lower on relevance, as they leave some part of the message unexpressed, though they will still score higher than completely irrelevant constructions.

Item-in-construction frequency is simply the frequency with which each item (most importantly, for this example, the verb) has appeared in each of the competing constructions. Items in the message (e.g., *pull*) boost the activation of those constructions in which they have previously appeared, with a strength proportional to the frequency of previous occurrence (e.g., Gahl, Jurafsky, & Roland, 2004; Trueswell, Tanenhaus, & Kello, 1993).

Thus, under the present account, pre-emption arises due to the effect of item-in-construction frequency operating over competing constructions that score similarly, and highly, on relevance. For the example message "Marge caused Homer to have a box by pulling a box towards him", the two constructions that score highest on relevance are the PO- and DO-dative. Thus, the winner will be determined by (in addition to semantic/ morphophonological fit) item-in-construction frequency (i.e., the number of times that *pull* has appeared in each of these constructions; i.e., pre-emption).

With regard to entrenchment, it is important to emphasise that it is not *only* the two most relevant constructions that compete for the right to express the message; and hence for which fit and item-in-construction frequency are relevant. All constructions that are at least somewhat relevant—including, for the present example, the simple transitive and intransitive constructions—will receive some degree of activation on the basis of their relevance to the message. This means that item-in-construction frequency is important, not only for the two most relevant constructions (here, the PO- and DO-dative), but also for other constructions in which the item has appeared, including the transitive and intransitive. Thus, the account predicts an entrenchment effect arising from occurrences of the verb (here *pull*) in constructions such as the transitive and intransitive. Thus, under the present account, entrenchment and pre-emption are not separate mechanisms as such. Rather, all constructions compete for the right to express the speaker's message. The effect that has historically been called "pre-emption" describes the effect exerted by the construction most relevant for the speaker's meaning; "entrenchment" the effect exerted by those constructions whose meaning is more tangential. Whether the closer ("pre-emption") or more distant ("entrenchment") constructions have more influence may well depend on the relative frequency of the two in the input, though this is a question for future research.

It is important to acknowledge, however, that—at present—this account does not make precise quantitative predictions regarding human grammaticality judgement data, as the relative influence of—and interaction between—the factors of (1) construction frequency, (2) semantic/morphophonological fit, (3) relevance, and (4) item-in-construction frequency has yet to be spelled out explicitly. In order to do so, it will probably be necessary to implement the account as a computational model. Useful starting points may be the Bayesian simulations of Alishahi and Stevenson (2008), Perfors et al. (2010) and Parisien and Stevenson (2011), the latter of which

UNCORRECTED PROOF

modelled the pattern of judgements observed by Ambridge et al. (2011). The connectionist Dual-Path model of Chang, Dell, and Bock (2006) has also been shown to simulate certain aspects of the retreat from overgeneralisation. In future work, we aim to compare Bayesian and connectionist models in terms of their ability to explain the pattern of grammaticality judgement data obtained from both adults and children, over a variety of different constructions.

Until the present verbal account is implemented as a computational model, it is limited with regard to the extent to which it makes testable quantitative developmental predictions. However, even in its present state, the account would seem to make three developmental predictions that potentially differentiate it from its rivals. All three predictions stem from the assumption that children's overgeneralisation errors reflect the use of verbs in construction slots with which they are less than optimally compatible (for adults), due to immature knowledge of the semantic properties of verbs and/or construction slots. Errors cease as children refine their knowledge of these semantic properties of verbs and slots.

The first prediction is that the extent to which children's grammaticality judgements can be predicted by adult-rated fine-grained semantic properties should increase with development. Although the present study provided some support for this prediction, a potential difficulty here is that such an effect could be observed even in the absence of a growing knowledge of semantics, if children's judgements become increasingly less noisy with age. Although it is almost certainly the case that younger children's judgements are more noisy than those of older children and adults, this cannot explain the entire developmental pattern observed, and, in particular, the finding that broad-range semantic properties explained more variance for 5- to 6-year-olds than for either of the older groups. Thus, the claim that children's knowledge of verb and construction semantics becomes increasingly fine-grained with development would seem to enjoy some preliminary support.

This does not, of course, constitute evidence that errors are caused by incomplete knowledge of verb and construction semantics, or that errors disappear as this knowledge is acquired. Thus, the second developmental prediction of the present account is that only children who display non-adult-like knowledge of the semantic properties of a particular verb will produce overgeneralisation errors with that verb (or rate such errors as acceptable). Preliminary support for this prediction, for a single pair of verbs, comes from the study of Gropen, Pinker, Hollander, and Goldberg (1991). These authors found that children who produced locative overgeneralisation errors with *fill* (e.g., **Lisa filled water into the cup*) judged pouring events where a glass ended up only $^3/_4$ 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 that they produced such errors because verbs with this meaning are indeed compatible with this construction (e.g., *Lisa poured water into the cup*). Future studies should investigate the relationship between children's knowledge of verb semantics and their propensity to produce, or accept, overgeneralisation errors with a wider range of verbs and—amongst others—the PO- and DO-dative constructions.

The final prediction of the present account is that children will produce errors as a function of non-adult-like knowledge of the semantic properties of particular construction slots (e.g., that the VERB slot in the DO-dative admits only verbs that denote possession transfer). A paradigm that may be suitable for testing this prediction is the novel-construction-learning paradigm of Goldberg and colleagues (Boyd, Gottschalk, & Goldberg, 2009; Casenhiser & Goldberg, 2005; Wonnacott, Boyd, Thompson, & Goldberg, 2012). These studies involve teaching children novel

UNCORRECTED PROOF

argument-structure constructions characterised by a novel word order (and, in some cases, novel morphemes) associated with a particular meaning, such as appearance. The present account predicts that only children who fail to appreciate that the verb slot requires a filler that is semantically compatible with the notion of appearance will produce errors by generalising incompatible verbs into the construction.

In conclusion, whether or not the account that we have outlined here is supported by future research, the results of the present study demonstrate that any successful account of the retreat from argument-structure overgeneralisation errors will include a role for pre-emption/entrenchment and verb semantics (and perhaps—for the dative constructions—morphophonology). Our goal has been to outline one possible account that combines these factors. It is to be hoped that future studies designed to test the predictions of this account, perhaps incorporating a modelling component, will bring the field closer to an understanding of how children retreat from, or avoid, overgeneralisation errors.

Manuscript received 3 February 2012
Revised manuscript received 26 September 2012
First published online month/year

UNCORRECTED PROOF

REFERENCES

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science*, *32*(5), 789–834.

Ambridge, B. (2011). Paradigms for assessing children's knowledge of syntax and morphology. In E. Hoff (Ed.), *Guide to research methods in child language* (pp. 113–132). London: Blackwell-Wiley.

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches.* Cambridge: Cambridge University Press.

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*, 303–323.

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, *123*, 260–279.

Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (2012). The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language*, *88*, 45–81.

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., & 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.

Arnon, I., & Clark, E. V. (2011). Why *Brush Your Teeth* is better than *Teeth*: Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*, 107–129.

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

Barr, D. J., Levy, R., Scheppers, C., & Tilley, H. J. (submitted). Random effects structure for confirmatory hypothesis testing: Keep it maximal. **AQ15**

AQ16 Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science*, *33*(5), 752–793.

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. **AQ17**

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.

AQ18 Bresnan, J. (2007a). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential case*. Berlin: Mouton de Gruyter.

Bresnan, J. (2007b). A few lessons from typology. *Linguistic Typology*, *11*(1), 297–306.

Bresnan, J., Cueni, A., Nikitina, T. V., & Baayen, H. R. (2007). Predicting the dative alternation. In G. Boume, I. Kramer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., & Nikitina, T. (2007). The gradience of the dative alternation. In L. Uyechi & L. Hee Wee (Eds.), *Reality exploration and discovery: Pattern interaction in language and life*. Stanford: CSLI Publications.
AQ19 http://lfg-stanford.edu/bresnan/download.html

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, *75*(4), 720–738.

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.

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

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied Psycholinguistics*, *22*(2), 253–267.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

Chater, N., & Vitanyi, P. (2007). 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, *51*(3), 135–163.

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, *30*(3), 637–669.

Clark, E. V., & Bernicot, J. (2008). Repetition as ratification: How parents and children place information in common ground. *Journal of Child Language*, *35*(2), 349–371.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, *55*, 767–811.

AQ20 Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

de Marneffe, M-.C., Grimma, S., Arnon, I., Kirby, S., & Bresnan, J. (2011). A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes*, *27*(1), 25–61.

Dowman, M. (2000). Addressing the learnability of verb subcategorizations with Bayesian inference. In L. Gleitman & A. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 107–112). Austin, TX: Cognitive Science Society.

Edelman, S., & Christiansen, M. (2003). How seriously should we take Minimalist syntax? *Trends in Cognitive Sciences*, *7*, 60–61.

Felbaum, C. (2005). Examining the constraints on the benefactive alternation by using the World Wide Web as
AQ21 a corpus. In M. Reis & S. Kepser (Eds.), *Evidence in linguistics: Empirical, theoretical and computational perspectives*. Berlin: Mouton de Gruyter.

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers*, *36*, 432–443.

Gahl, S., & Yu, A. C. L. (2006). Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review*, *23*, 213–216.

Gibson, E., & Fedorenko, E. (2011). The need for quantitative methods in syntax and semantics research.
AQ22 *Language & Cognitive Processes*. doi:10.1080/01690965.2010.515080

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

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*, 131–154.

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

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language*, *18*(1), 115–151.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, *65*(2), 203–257.

UNCORRECTED PROOF

AQ23 Hahn, U., & Oaksford, M. (2008). Inference from absence in language and thought. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind*. Oxford: Oxford University Press.

AQ24 Hsu, A. (2009). Differential use of implicit negative evidence in generative and discriminative language learning. *Advances in Neural Information Processing Systems*, *22*.

Hsu, A., & Chater, N. (2010). The logical problem of language acquisition goes probabilistic: No negative evidence as a window on language acquisition. *Cognitive Science*, *34*, 972–1016.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: *The What's X doing Y?* construction. *Language*, *75*, 1.

AQ25 Krifka, M. (2004). Semantic and pragmatic conditions for the dative alternation. *Korean Journal of English Language and Linguistics*, *4*, 1–32.

Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford, CA: Stanford University 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.

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

AQ26 Lidz, J., Gleitman, H., & Gleitman, L. R. (2004). Kidz in the 'hood': Syntactic bootstrapping and the mental lexicon. In G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon*. Cambridge, MA: MIT Press.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, *46*(1), 53–85.

McNeill, D. (1966). Developmental psycholinguistics. In F. Smith & G. A. Miller (Eds.), *The genesis of language: A psycholinguistic approach* (pp. 15–84). Cambridge, MA: MIT Press.

Naigles, L., Fowler, A., & Helm, A. (1992). Developmental changes in the construction of verb meanings. *Cognitive Development*, *7*, 403–427.

AQ27 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.

Oehrle, R. T. (1976). *The grammatical status of the English dative alternation* (Unpublished doctoral dissertation), MIT, Cambridge, MA.

Onnis, L., Roberts, M., & Chater, N. (2002). Simplicity: A cure for overregularizations in language acquisition? In W. Gray & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 720–725). Austin, TX: Cognitive Science Society.

Parisien, C., & Stevenson, S. (2011, July). *Generalizing between form and meaning using learned verb classes.* Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Boston, MA.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*, 306–338.

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

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

AQ28 Ross, J. R. (1967). *Constraints on variables in syntax* (doctoral dissertation). MIT [Published as *Infinite Syntax!*. 1986. Norwood, NJ: Ablex Publishing Corporation].

Rowland, C. F., & Noble, C. (2010). The role of syntactic structure in children's sentence comprehension: Evidence from the dative. *Language Learning and Development*, *7*(1), 55–75.

Rowland, C. F., & Theakston, A. L. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 2: The modals and auxiliary DO. *Journal of Speech Language and Hearing Research*, *52*(6), 1471–1492. doi:10.1044/1092-4388(2009/08-0037a)

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

Saxton, M., Backley, P., & Gallaway, C. (2005). Negative input for grammatical errors: Effects after a lag of 12 weeks. *Journal of Child Language*, *32*(3), 643–672.

Schutze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology.* Chicago, IL: University of Chicago Press.

Snyder, W., & Stromswold, K. (1997). The structure and acquisition of English dative constructions. *Linguistic Inquiry*, *28*(2), 281–317.

Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, *115*, 1497–1524.

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.

Strapp, C. M., Bleakney, D. M., Helmick, A. L., & Tonkovich, H. M. (2008). Developmental differences in the effects of negative and positive evidence. *First Language*, *28*(1), 35–53.

UNCORRECTED PROOF

{PLCP}articles/PLCP738300/PLCP_A_738300_O.3d[x] 15-10-2012 19:30:26

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*(3), 528–553.

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

Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Novel construction learning in 5 year olds: **AQ29** The role of input and age.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209.

UNCORRECTED PROOF