Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition

Katherine E. Twomey, Franklin Chang & Ben Ambridge
University of Liverpool, UK

Address correspondence to Franklin Chang, University of Liverpool, Department of Psychological Sciences, Eleanor Rathbone Building, Bedford Street South, Liverpool, L69 7ZA, UK.
Email: franklin.chang@liverpool.ac.uk. Telephone: +44-151-794-1110

Abstract

Children overgeneralise verbs to ungrammatical structures early in acquisition, but retreat from these overgeneralisations as they learn semantic verb classes. In a large corpus of English locative utterances (e.g., *the woman sprayed water onto the wall/wall with water*), we found structural biases which changed over development and which could explain overgeneralisation behaviour. Children and adults had similar verb classes and a correspondence analysis suggested that lexical distributional regularities in the adult input could help to explain the acquisition of these classes. A connectionist model provided an explicit account of how structural biases could be learned over development and how these biases could be reduced by learning verb classes from distributional regularities.

1.1 Introduction.

To learn a language, children must learn how to link verbs to abstract grammatical structures. For example, they must learn that the locative verb *fill* can occur in sentence structures such as *I filled the salt shaker with salt*. However, young children also overgeneralise verbs to structures in which they are ungrammatical (e.g., E(5;0) *Can I fill some salt into the bear* [bear-shaped salt shaker], Bowerman, 1982; see also Ambridge, Pine, & Rowland, 2012; Pinker, 1989). These overgeneralisations show that children understand something about verb meanings and sentence structures but have not fully learned the appropriate pairing of verbs and structures. Over time, children learn to constrain their choice of structure and begin to "retreat" from overgeneralisation (Ambridge, Pine, Rowland, Jones, & Clark, 2009; Pinker, 1989). This paper examines the nature of the mechanisms that support this retreat in the context of the acquisition of the English locative.

The English locative alternation consists of two structures: the location-theme structure (*LT structure*, e.g., *the woman sprayed the wall with water*), and the theme-location structure (*TL structure*, e.g., *the woman sprayed water onto the wall*). These structures provide alternative ways of conveying a meaning involving the thematic roles AGENT (e.g., *woman*), LOCATION (usually a surface or a location, e.g., *wall*) and THEME (usually a liquid or an object, e.g., *water*). The difference between the two structures arises from the mapping of roles to structural positions: the LT structure places the LOCATION noun in object position after the verb, followed by the THEME noun in a prepositional phrase, while the TL structure places the THEME noun in object position, followed by the LOCATION noun in a prepositional phrase.

Work on the locative alternation has been shaped by Pinker's (1989) broad/narrow range rule account of overgeneralisation and retreat. He argued that children's early understanding of how locative verbs map to these structures involves innate "broad range" linking rules (Gropen, Pinker, Hollander, & Goldberg, 1991a). On this account, the broad range rule for the locative alternation links the two possible construals of a locative action as in (1) and (2).

1) X causes Y to change state by means of moving Z to Y.

2) X causes Y to move into/onto Z.

In these construals, X encodes the animate entity which carries out the action, Y labels the most affected entity and Z labels the remaining entity. Children must learn that the English LT structure – which focuses on the LOCATION (Y) change of state – maps to construal (1) (e.g., *the man sprayed the wall with water*) and that the TL structure – which focuses on the motion of the THEME (Y) – maps to construal (2) (e.g., *the man sprayed the water onto the wall*). The fact that these construals are linked by innate linking rules in Pinker's theory explains why children generalise verbs heard in one structure to the other structure early in development (Gropen, Pinker, Hollander & Goldberg, 1989), producing overgeneralisation errors in which children use verbs in structures that are not licensed by the adult language. For example, **I'm going to cover a screen over me* (E(4;5); Bowerman, 1982) is ungrammatical in adult speech, and results from the overgeneralisation of an LT-biased verb into the TL structure.

As they become more experienced language users, children learn to constrain such errors. Pinker (1989) links this retreat from overgeneralisation to the acquisition of "narrow range" rules, which link semantic verb classes (derived from classes developed by Levin, 1985; 1993) to particular structures. For example, *cover* and *coat* both refer to an action where a LOCATION is obscured by a THEME. Verbs in this class appear only in the LT structure (e.g. *the woman coated the car with paint* vs. **the woman coated paint onto the car*). Other verbs like *pour* and *spill* involve actions where the THEME is a liquid that flows into some LOCATION in a certain manner. Verbs in this class appear only in the TL structure (e.g., *the woman spilled water onto the floor* vs. **the woman spilled the floor with water*). Finally, other verbs like *spray* and *squirt* involve actions where the THEME is a liquid which both moves in a certain manner and affects the LOCATION in a certain way. Verbs in these "alternating" classes can appear in either structure (e.g., *the woman sprayed the wall with water*; *the woman sprayed water onto the wall*). As narrow range rules specify the correct structural properties of these verbs, their acquisition supports the retreat from overgeneralisation and the development of adult-like verb-structure mappings.

Pinker's theory introduced the idea that different kinds of semantics explain changes in the locative alternation across development, and specifically that the retreat from overgeneralisation involves the acquisition of semantic verb classes. In this paper, we contrast two accounts of how these verb classes are acquired, which we call *situational* and *distributional* accounts. The following paragraphs describe the nature of the semantics assumed by each account, and the different cues that signal syntactic distinctions. Because we define the problem of learning the locative alternation in different semantic and syntactic terms from previous theories of language acquisition, we conclude this introduction with a discussion of how our account relates to these existing theories.

One broad approach derives from the view that semantic knowledge is learned from cross-situational regularities between linguistic forms and observed events (St Augustine, 397/2001; Pinker, 1994; Smith & Yu, 2008). On this *situational* account, locative verb classes are learned from the non-linguistic features of the situations in which locative verbs are heard. For example, a child might hear *coat* whilst seeing a surface changing state as it becomes obscured by a liquid. Because the LOCATION is saliently changed in this situation, the child infers that *coat* (and verbs that share its container-state-change semantics) must be used with the LT structure, which places the LOCATION before the THEME. Gropen and colleagues (1991a) offered evidence for the situational account in a verb learning experiment, in which adult and child participants were familiarised with two action/verb pairs. Each pair involved the same LOCATION item and THEME item (e.g., a cloth and a matchbox, E1) and each could be described with a locative sentence. However, one action highlighted the manner in which the THEME moved, while the other highlighted the eventual endstate of the LOCATION. At test, the experimenter acted out the action and asked participants to describe it using the trained verb. Participants generalised the novel verb to the LT or TL structure based on the salience of the LOCATION and the THEME. Although this study showed that situational meaning can influence structural choice, it did not show that situational meaning assigns verb class: structural choice could be directly signalled by the relative salience of the LOCATION and THEME roles in the test situation without participants having learned verb class information during the familiarisation phase.

In thinking about the situational account, it is important to distinguish different types of situational information. One type of information is thematic roles (Gruber, 1965; Jackendoff, 1983), for example AGENT, LOCATION, and THEME. Although these roles license a locative, they are compatible with both the LT and TL structure. Another type of information, which we will call lexical semantics, refers to the semantics that is inherent to the verb across different syntactic frames (Pinker, 1994, calls this *root meaning*). For example, *cover* has a similar meaning in both transitives (e.g., *I covered the dishes*) or locatives (e.g., *I covered the dishes with water*). Because lexical semantics does not always determine a verb's structural preferences, we distinguish lexical semantics from syntax-related semantics such as

change of state (*frame meaning*; Pinker, 1994), which includes verb class semantics (Hare, McRae, & Elman, 2003). It is not trivial to explain how children might isolate these different situational semantic components and link them in appropriate ways to structural choices (Gillette, Gleitman, Gleitman, & Lederer, 1999).

If verb-structure mappings are shaped by the situational semantics in human experience, there should be similar mappings cross-linguistically. In fact, there is considerable variation. For example, the Korean locative alternation uses two structures to highlight LOCATIONS and THEMES (Kim, Landau, & Phillips, 1999), but unlike in English, the Korean verb *fill* can occur in both LT and TL structures (e.g., *John*-nom *water*-acc *cup*-loc *fill*-past-decl vs. *John*-nom *cup*-acc *water*-with *fill*-past-decl). Kim et al. (1999) found that while English mothers never used the English TL structure to describe a video of a filling event, Korean mothers used the Korean TL structure 57% of the time. Since the video was the same for both groups of mothers, the difference between Korean and English mothers' choice of LT or TL structure suggests that situational verb meaning may not fully explain structural choice. Similarly, Hunter (2008) describes cross-linguistic variation in the structural biases of locative verbs; for example, *splash* can alternate in English, but only appears in the LT structure in Greek, and only in the TL structure in Hebrew. This variation is hard to explain if, as per the situational account, these verb-structure mappings are mainly constrained by universal situational regularities.

The difficulty of constraining verb class acquisition from situational information has motivated *distributional* accounts in which children learn verb classes from regularities in their linguistic input (Fisher, Gleitman & Gleitman, 1991; Gleitman, Cassidy, Nappa, Papafragou & Trueswell, 2005; Mintz, 2003; Mintz, Newport & Bever, 2002; Redington, Chater & Finch, 1998). One type of distributional learning is syntactic bootstrapping (Fisher, Gertner, Scott & Yuan, 2010; Gleitman, 1990; Landau & Gleitman, 1985), where verb-structure regularities are used to infer verb meaning. For example, if *cover* occurs frequently in the LT structure (e.g., *I covered the table with paper*) and never in the TL structure (e.g., **I covered paper onto the table*), then *cover* is assigned to a verb class that only allows the LT structure. Another approach to distributional learning is to assign syntax-relevant semantics using the words that co-occur with the verb (Dorr & Jones, 1996; Dumais & Landauer, 1997; Joanis, Stevenson & James, 2008; Resnik, 1996; Rohde, Gonnerman & Plaut, 2006; Riordan & Jones, 2011; Sun and Korhonen, 2009; Redington et al., 1998). For example, if a child hears the utterance *He is sloshing **paint** around*, the child classifies *slosh* with other verbs that take **paint** as an object (e.g., *the man poured **paint** into the bucket*; *the girl spilled **paint** on the table*), creating a verb class based on word distributional similarities. Models of this process cluster together verbs that co-occur with similar sets of words in a similarity space. Because this similarity space is often thought to be semantic in nature (e.g., Latent Semantic Analysis; Dumais & Landauer, 1997) these clusters of verbs approximate semantic verb classes that could be used to constrain structural choices.
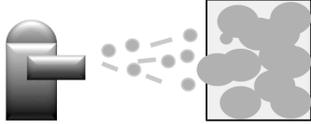
	Situational verb learning	Distributional verb learning
Linguistic input	*"Look, he's coating it!"*	*"coat the wall"* *"he coated the table with paint"* *"she's coating the canvas*
Situational input		None or inconsistent
Thematic roles	AGENT = robot LOCATION = wall THEME = paint	Occasionally inferable from utterance
Lexical semantics	*"coat"* refers to event where a liquid adheres to a surface	Lexical meaning of "coat" is occasionally inferable from utterance
Syntax-relevant semantics	LOCATION (wall) is more saliently changed than theme (PAINT) → *"coat"* is associated with LOCATION-biased semantic verb class	*"wall"*, *"table"*, *"canvas"* create similarity to other verbs with similar arguments → *"coat"* is associated with semantic verb class of words that have similar arguments (e.g., LOCATION-biased verbs like *"cover"*)

Test

Situational input	
Thematic roles	AGENT = robot LOCATION = canvas THEME = ink
Lexical semantics	Ink adheres to canvas → *"coat"* is an appropriate verb
Syntax-relevant semantics	Robot causes ink to move towards canvas; canvas is changed by ink; Ink and canvas are equally salient → Both LT and TL locative structure are possible descriptions
Linguistic output	Location-Theme structure: *"the robot coated the canvas with ink"* Theme-Location structure: *"the robot coated ink onto the canvas* If child learned verb class for "coat", then only LT utterance is acceptable

Figure 1. Distributional and situational accounts of children's acquisition of the English locative construction

To understand the differences between the situational and distributional accounts, it is important to first distinguish between the learning and testing situations (Figure 1), a difference which is not always clearly isolated in existing theories. The situational and distributional accounts only apply to verb class learning (also called familiarisation, training). At test – that is, when a child is asked to describe an event with more than one possible construal – all theories and models assume that locative production is supported by semantic and thematic role information. First, the speaker must have lexical semantics for the event they wish to describe, otherwise they will select the wrong verb (e.g., saying *eat* instead of *coat*). Second, the speaker must identify the AGENT, LOCATION, and THEME in the event that they want to convey. If these roles are not assigned, then a locative structure is inappropriate. Furthermore, it is thought that the relative salience of arguments can determine structural choices independently of the verb (Bock 1982; Gleitman, January, Nappa & Trueswell, 2007; Goldberg, 1995). Indeed, Gropen et al. (1991a) demonstrated that changes in the salience of LOCATION and THEME at test did change

the structural choices that children made with novel locative verbs. Thus, the LOCATION and THEME must be equally salient at test, in order to be sure that the child is producing the LT or TL locative structure based on a learned verb class rather than situational test cues. Although the test situation provides children with a range of semantic information that affects their choice of locative verb and structures, they must nonetheless acquire verb classes via either situational or distributional learning in order to know whether a particular verb is LT-biased, TL-biased, or alternating (e.g., *coat* is LT-biased, *spray* is alternating).

Thus, children may use all available information when producing locative structures at test, however situational and distributional theories provide different accounts about how learning takes place before the test situation. In the situational account verb bias is inferred from the salience of roles in an event. For example in Figure 1 (top left), if the wall (LOCATION) is more saliently changed than the paint (THEME), then *coat* belongs to a location-biased verb class. In the distributional account, children assign verbs to verb classes based on linguistic similarity: *coat* occurs with similar words to other LT-biased verbs like *cover* (Figure 1, top right). Thus, although both these theories assume that verbs are assigned to classes that predict their structural biases, they differ in the information that is used.

Another important difference between the situational and distributional accounts is the likelihood of a learning episode, and this can have important consequences for low-frequency verbs. For example, the low frequency verb *litter* is only acceptable in the LT structure (e.g. *I littered the table with papers*). On a strong situational account, a learning opportunity occurs when the word *litter* is heard in a situation where *littering* is visibly taking place. To correctly classify *litter* as a non-alternating LT-biased verb, some aspect of that situation must make the table more salient than the papers. In contrast, on a strong distributional account a learning opportunity occurs when a child hears any utterance with a verb usage of the word *litter*. The child can use the fact that *litter* occurs with a noun (e.g., *table*) that has appeared with another verb such as *cover* (e.g., *I littered the **table**, I covered the **table***) to classify those verbs as more similar in their structural preferences. Because simply hearing an utterance is more likely than hearing that utterance at the same time as seeing a depiction of the action with clear LOCATION/THEME salience cues, learning episodes for the distributional account should be more frequent than those for the situational account. Thus, according to situational theories, there may be low-frequency locative verbs (e.g., *bestrew*, *cultivate*, *dribble*, *encrust*, *festoon*, *imbue*, *lard*, *mottle*, *replenish*, *shroud*, *wad*; Levin, 1993) which are never encountered in an appropriate situational context and hence would not have a verb class.

While the situational and distributional theories discussed here are similar to other theories of language acquisition, they differ in some important ways. As noted, the situational account emphasises the role of semantics in verb class acquisition, which has been interpreted as a type of bootstrapping. It is important to clarify, however, that the situational account as defined here is *not* a version of Pinker's "semantic bootstrapping", which is a mechanism by which semantics is used to generate hypotheses for learning syntactic rules (Pinker, 1994). Pinker clearly distinguishes semantic bootstrapping from his semantic structure theory of linking (broad/narrow rules) in the conclusions of his locative paper (Gropen et al., 1991a), whereas our situational account relates to how children learn the semantic classes that support the narrow range rules. The distributional account is also different from syntactic bootstrapping theories, which argues that children use abstract syntactic frames to learn about verbs (Gleitman, 1990). In terms of the locative structure, the frame used with locative verbs consists of the sequence of syntactic categories NP V NP PP. If syntactic bootstrapping requires that frames consist solely of abstract syntactic categories (without lexical or thematic role distinctions), then children could not learn the locative alternation using this information, because the LT and TL structures both use the same NP V NP PP sequence of syntactic categories. Therefore for syntactic bootstrapping to explain acquisition of locative verb classes, children's frames must employ lexical elements or thematic roles to distinguish the two structures. Because locative acquisition takes place relatively late in development, children may have learned relevant syntactic frames by the time they are learning about the properties of locative verbs; indeed, computational models show that such syntactic frames can be used to learn about properties of locative verbs (Niyogi, 2002).

In this paper, we use a range of methods to better understand locative verb acquisition. Previous work has looked mostly at overgeneralisation errors in language development. Here, we examined children's normal locative use over development in a large corpus analysis of both high and low frequency verbs. Because it is difficult to characterise the situational input that was available when these verbs were learned, we analysed the degree to which distributional learning can account for verb class acquisition using several correspondence analyses of child-directed speech (Section 2). Finally, we used a connectionist model of language acquisition to better understand how verb classes can be acquired (Section 3).

2.1 A Corpus Analysis of the English Locative Construction

Although the English locative has been the focus of a handful of corpus, diary, empirical, elicitation and grammaticality judgment studies (Ambridge et al., 2012; Bidgood, Ambridge, Pine, & Rowland, 2014; Bowerman, 1982; Gropen et al., 1991a, 1991b; Laffut & Davidse, 2002), to our knowledge there exists no in-depth examination of the patterns of locative verbs in the language children hear and use. Of the few previous corpus studies of the English locative, some examine a small set of verbs, others use a small set of corpora, and others use only adult or non-naturalistic data. Laffut and Davidse (2002), for example, examined eight locative verbs from the adult COBUILD corpus of written text; Iwata (2008) counted occurrences of five locative verbs in the British National Corpus of adult written text and speech; and Bowerman (1982) presented examples of

locative overgeneralisations based on a corpus of speech from five children from approximately two to seven years old. In contrast to these smaller studies, we examine the behaviour of a wide range of locative verbs in the 14 UK English corpora in the CHILDES child language database (MacWhinney, 2000). We extracted all utterances in the UK corpora in CHILDES (Cruttenden, 1978; Fletcher & Garman, 1998; Forrester, 2002; Gathercole, 1986; Henry, 1995; Howe, 1981; Johnson, 1986; Korman, 1992; Lieven, Salomo, & Tomasello, 2009; Rowland & Fletcher, 2006; Theakston, Lieven, Pine & Rowland, 2001; Tommerdahl, 2009; Wells, 1981; Wilson & Henry, 1998; Wooten, 1984) for the 140 locative verbs for which Ambridge et al. (2012) collected grammaticality ratings. We created the *full corpus* of 38,231 utterances by searching for all possible forms of each of the 140 locative verbs (e.g., *stick, sticking, stuck*), and found tokens of 103 forms (see Appendix A for details). Utterances came from 1,975 transcribed files of recorded sessions between adults and children ranging from four weeks to seven-and-a-half years old.

Next, we created a *handcoded corpus,* which we coded for various syntactic structures. Our goal was to have a large sample of locative verbs to allow us to characterise their structural preferences, so we selected utterances from the full corpus which had been tagged as verbs or participles on the CHILDES MOR line. We then coded each utterance for prepositions anywhere post-verbally (*with*, indicating a candidate LT locative; *into*, *onto*, or *over*, indicating a candidate TL locative; or no preposition, indicating some other construction). We counted the number of utterances separately for each verb in these *LT candidate, TL candidate,* or *Other* categories for adults and children. When this count was less than 50, we coded all utterances in that category. When it was more than 50, we coded a random sample of 50 utterances. Because there were very few tagged utterances for children over 40 months, we extracted a further 20% of child utterances and 5% of adult utterances from the untagged data (i.e., no MOR line) for this age group, resulting in 2,685 utterances in our handcoded corpus. Each utterance was coded individually, with no preceding or succeeding conversational context. In addition to coding LT and TL structures we also coded transitive utterances, which were distinguished by plausibility of the post-verbal arguments as LOCATION or a THEME (Table 1). Transitives with post-verbal LOCATION nouns (e.g., *the man sprayed the wall*) were labelled as *L transitives* and transitives with post-verbal THEME nouns (e.g., *the man sprayed the water*) were labelled as *T transitives*. Finally, ambiguous transitives with post-verbal pronouns (e.g. *the man sprayed it*) and intransitives were coded separately. A further 931 non-verb (miscoded on the MOR line or from untagged data; e.g., nouns: *Give me the brush*; adjectives: *I think it's a bit smeary*) or ambiguous utterances were excluded from subsequent analyses. An additional 10% of responses were coded by a second coder. Intercoder reliability was substantial (Cohen's kappa = 0.71; Landis & Koch, 1977).

2.1.1 Verb Classes and Structural Biases in Children and Adults
To begin to understand how children acquire locative verb classes, we examined whether children's locative verb use in the coded corpus exhibited structural biases similar to those in adults. As the data for some verbs were sparse, we focused on the 15 most frequent verbs and calculated the proportion of location-biased structures (i.e., proportion L transitives and LT locatives out of all transitives and locatives; Figure 2). Overall, children's locative verb use reflected adults' locative verb use. The verbs *cover, fill,* and *brush* appeared almost exclusively in L transitives and LT locatives in both adult and child utterances. Similarly, *stick, pour,* and *dump* occurred almost exclusively in T transitives and TL locatives in both adult and child utterances. The verbs *rub, spill* and *splash* alternated between the two types of structure, again in both adult and child utterances. This demonstrates that for high frequency verbs, structural biases in parental input can model the biases that children learn.
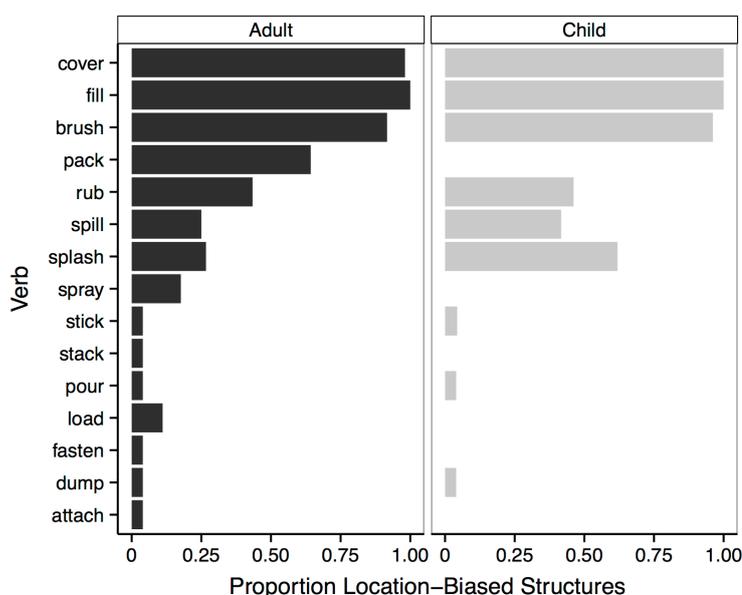

Figure 2. Proportion L transitive and LT locative structures for the 15 most frequent verbs in locative constructions separated by speaker

Table 1. Construction frequencies.

Construction type	Example	Frequency	
		Adult	Child
LT locative	brushing me with it?	93	7
TL locative	spread it on your biscuit	183	49
L transitive	fill the carriage	190	81
T transitive	you dump the lady's toys	231	63
Ambiguous transitive	brushed it	410	198
Intransitive	it just wound up	153	96

2.1.2 The Developmental Trajectory of Structural Choice in the English Locative

Overgeneralisations of locative structures are rare events which illustrate children's ability to extrapolate beyond their input. It has been suggested that these overgeneralisations arise from biases toward particular structures (e.g., one structure is more robust, or one structure is derived from another, more basic structure; Abbot-Smith & Behrens, 2006; Bowerman, 1982; Pinker, 1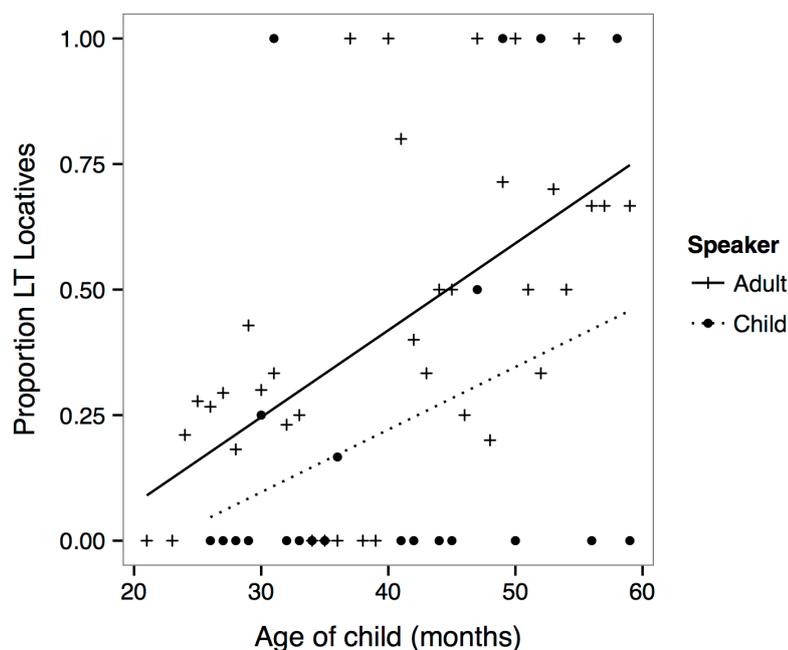989). If so, these biases should also be evident in children's choice of structure, and in this case, the choice between LT and TL locative structures. To examine whether our data provided evidence for such structural biases in development, we calculated the proportion of LT locative structures (out of all locatives) for children and adults at each age in months in the corpora (only 33 verbs were used in the locative). Because the 565 transitive uses of locative verbs outnumbered the 332 locative uses, we also calculated the proportion of L transitive structures out of all transitives (40 verbs were used in the transitive, 26 were in both).

Figure 3 displays the proportion of LT locative structures (e.g. *the woman sprayed the wall with water*), and Figure 4 displays the proportion of L transitive structures (e.g., *the woman sprayed the wall*), both for adults and children between 20 and 60 months old in the handcoded corpus. When verbs appeared in locatives early in development, they occurred most frequently in the TL structure (Figure 3). In contrast, transitive use remained consistent over time, with around 50% of verbs occurring in the L structure (Figure 4). A linear model was fitted to location-biased structures produced (L or LT = 1, T or TL = 0) with structure type (transitive, locative), child age (months) and speaker group (adult, child) transitive production (beta = -0.91, $t(116) = 3.06$, $p = .003$), pointing to a *TL bias* in locative production. An interaction between child age and structure type (beta = 0.02, $t(116) = 3.12$, $p = .002$) indicated an increase in LT production over development. This analysis reveals that transitive and locative uses of locative verbs differ in their respective proportion of L and LT structures (in which the LOCATION role occurs directly after the verb).

In Figure 3, adults' use of full locatives seems to mirror the changes in children's use of locatives over development. Since adults should be able to use both LT and TL structures, the asymmetry in the adult utterances must be explained. One possibility is that it is due to structural priming (Bock, 1986), but this would mean that adult use of LT depended on being primed by a child LT structure. Another possibility is that adults tend to repeat the same verb-structure pairs that the children use. To test this, we extracted the 207 separate files that contained full locatives to examine how often children use locatives with a particular verb in the same file as an adult use of that verb. We found that only 3.38% (7 files) included both

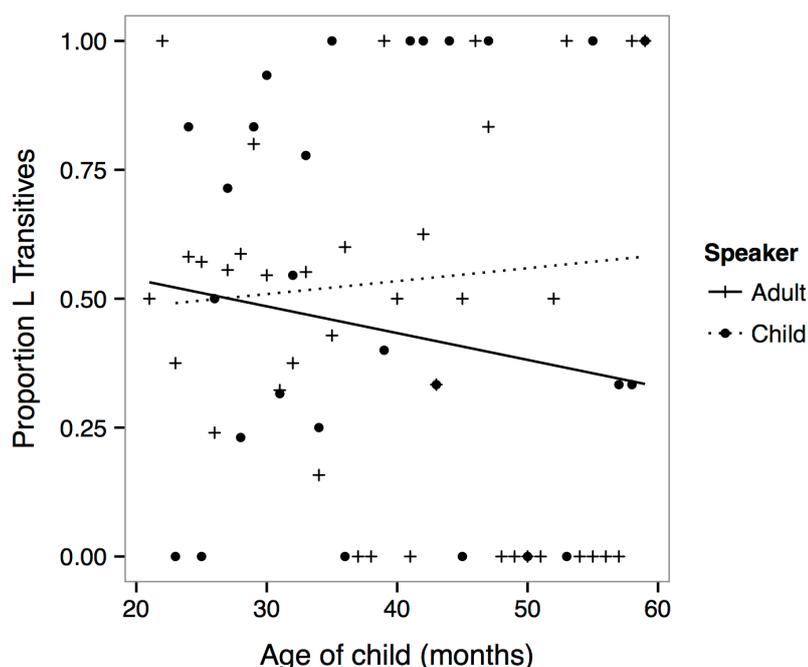*Figure 3*. Proportion LT locatives out of all coded locative utterances in the corpus data separated by age and speaker

Figure 4. Proportion L transitives out of all coded transitive utterances in the corpus data separated by age and speaker

child and adult uses of the same verb in a locative structure. Thus, it is unlikely that the developmental effect is due to imitation or priming. A more likely explanation is that adults used more TL structures early in children's acquisition as a general adjustment to make their utterances comprehensible to children (motherese; Bohannon & Marquis, 1977; Gleitman, Newport & Gleitman, 1984). Importantly, the early TL bias shown in Figure 3 suggests that parents' full locatives information for children to learn LT-biased verbs. Instead, the consistency of transitive use over development (Figure 4) suggests that transitive uses of locative verbs may be a more stable basis for learning verbs' structural biases than locative uses of locative verbs. Furthermore, transitive uses are more frequent than locative uses: while 16 verbs appeared more frequently in the locative than in the transitive, 30 verbs appeared more often in the transitive than in the locative. Given the TL bias and sparseness of locatives in the input, it would make sense for a theory of locative verb class acquisition to take advantage of transitive uses of these verbs.

To examine whether transitive use is predictive of locative uses, we used mean adult ratings of locative verbs' LT and TL biases taken from Ambridge et al. (2012) and Bidgood et al. (2014) to create a graded LT preference rating measure for each verb (100% LT preference = 1, 100% TL preference = 0). We correlated this with the proportion of LT and L structures in the adult utterances in the corpora. There was a significant correlation between LT rating and adult LT distribution ($r = .53$, $t(30) = 3.40$, $p = .002$), indicating that the LT bias of verbs in locative structures observed in our child-directed speech data matched adult ratings of those verbs' bias towards LT structures. There was also a significant correlation between LT rating and adult L distribution in the corpus ($r = .32$, $t(38) = 2.11$, $p = .041$), indicating that the biases of verbs in *transitive* structures in our data also reflect adults' locative preference ratings. These correlations suggest that distributional regularities in transitives could be used to learn locative verb biases.

As noted above, young children significantly preferred post-verbal THEMES in the locative construction (e.g., *the woman sprayed water into the bucket*) but not in the transitive construction (e.g., *the woman sprayed water*). In children under 30 months old, 75% of the locative utterances were TL structures, but only 44% of the transitive utterances were T structures. The TL bias is mirrored in existing empirical, diary and corpus studies (Bidgood et al., 2014; Bowerman, 1982; Gropen et al., 1991a, 1991b; Laffut & Davidse, 2002). For example, Bidgood et al. (2014) found in a novel verb ratings task that five- to six-year-old children accepted novel verbs in the TL locative structure but not the LT locative structure. Further, Gropen et al. (1991b, E1) found that 3;4 - 4;5-year-old children placed the THEME in the post-verbal position 60% of the time even when describing an endstate event that emphasised the location and prompted with a location-biased question. Critically, the TL bias may be the basis for locative overgeneralisations: if a child hears an LT-biased verb in a LT structure (e.g., *I will cover the fireplace with a screen*), but they themselves have an overall TL bias, this broader bias could prompt them to use an LT-biased verb in a TL

structure. Bowerman (1982) found that 79% of her locative overgeneralisation errors were of exactly this type (e.g., E(4;5) "I'm going to cover a screen over me"). Thus, the TL bias for normal utterances in our corpus data could be a metric of the structural bias that creates these TL overgeneralisation errors, and the significant increase in LT use over development suggests a retreat from these structural biases. Later, we explore whether biases in normal production can explain errors using a computational model (Section 3).

2.2 A Corpus-Based Test of the Distributional Learning Hypothesis

The distributional account of locative verb class acquisition poses several as-yet unanswered questions: what kind of distributional regularities might children use to learn verb classes? Is there sufficient information in the linguistic input to support verb class acquisition? If so, can these verb classes predict whether adults will use an LT or a TL structure? To address these questions, we implemented the distributional account in a series of correspondence analyses (CA; Greenacre, 2007).

Our CAs clustered target verbs based on the words that they co-occurred with, generating multiple components akin to those produced by principal component analysis. The CAs created a low-dimensional similarity space that accounted for the variance in the higher dimensional space of word-to-word co-occurrences (a contingency table). For example, *spray* and *splash* might both be followed by *some water* and *the paint*, while *dappled* might be only followed by *some paint*. Because *spray* and *splash* share four post-verbal words, they would be classified as more similar to each other than to *dappled*, with which they share two post-verbal words.

We prepared the input for the CA in two stages. First, to ensure we were classifying locative verbs, we selected target verbs from the 140 verbs which had an LT rating (see Section 2.1) and which occurred in our hand-coded data at least 25 times in a locative structure (Sun & Korhonen, 2009). This is important because many locative verbs also occur in non-locative forms. For example, although *blanket* is a locative verb, it only occurred as a noun in our hand-coded data. In an analysis based on co-occurrence, it is possible that *blanket* and *cover*, for example, might be classified as similar if their noun uses shared adjacent words (e.g., *the cover on the bed*; *the blanket on the bed*). Although these noun uses could help support verb class assignment, we focused here on those verbs that occurred frequently in the locative structure in the full corpus to ensure that our analysis directly reflected the acquisition of locative verb classes. This generated a target set of 21 verbs.

Next, we generated our input data by extracting all utterances that contained any form of the word (e.g., first person present *cover*, third person singular *nails*, progressive aspect *rubbing*, past participle *dumped*, irregular *spilt*) for each of our 21 target verbs. We used the full dataset (not just the hand-coded subset) and discarded any hand-coded syntactic information, so that input consisted solely of the words of each utterance. Importantly, although we selected target verbs which occurred frequently as locative verbs, our CAs were based on the entire dataset. For example, *cover* was classified

using the words which co-occurred with both verb uses (e.g., *like Mummy covers you up*) and noun uses (e.g., *here's the covers for your dolly*). The input to these CAs therefore simulated how a child could learn verb classes from heterogeneous input where non-locative uses are not filtered out.

To investigate what kind of distributional information might be useful in learning verb classes, we constructed multiple CAs with different inputs and evaluated their ability to predict adult LT ratings (Ambridge et al., 2012). The first analysis (*Full* CA) clustered verbs based on all other words in the utterance. Because the locative alternation occurs post-verbally, the second analysis used all post-verbal words (*Post-verbal* CA). A third, *Preposition-only* CA used only post-verbal prepositions (*at, in, into, on, onto, over, with*), as these are specific cues for LT and TL forms. The fourth, *Two Post-verbal Words* CA used only the two post-verbal words because adjacent word statistics are commonly used in computational models (e.g., trigrams, Chang et al., 2008; Redington et al., 1998; Mintz et al., 2002). Because our corpus analysis identified transitives as a possible source of locative verb class information, our final CA (*Transitive* CA) explored whether there was sufficient information in post-verbal articles and nouns (that is, in transitive but not locative utterances) to correctly classify locative verbs (Levin & Rappaport Hovav, 1991). This analysis therefore used the same input as the Two Post-verbal Words CA, but excluded prepositions as well as any words following these prepositions. Finally, to examine whether children could learn locative verb classes via syntactic bootstrapping, in a sixth analysis we used proportion LT in the hand-coded input to predict LT rating (*Proportion LT* analysis), simulating a child learning verb classes from the occurrence of locative verbs in adult-like LT or TL structures.

To determine which of the five CAs best predicted our LT rating measure, we used the first six factor scores in each CA in a regression to predict LT rating. For each regression, we selected the factor which most reliably predicted LT Rating and submitted it to a further regression to obtain a measure of variance accounted for (R^2). As can be seen in , the Two Post-verbal Words analysis best accounted for LT rating: Factor 4 significantly

predicted LT Rating and explained 47% of the variance. Figure 5 depicts this relationship, showing that verbs rated as more grammatical in TL structures (e.g., *pour* and *spill*) were positive on Factor 4, and verbs rated as more grammatical in LT structures (e.g., *cover* and *fill*) were negative on Factor 4. Further, some verb clusters reflect Pinker's (1989) narrow semantic verb classes; for example, *spray* and *splash* were similar on Factor 4 and are both members of Pinker's *spray*-type class. In line with existing studies (e.g., Ambridge et al., 2012) not all classes generated by our data conform to these narrow verb classes. For example, *spread* and *spill* were close on Factor 4, despite being members of different classes according to Pinker's classification. As well as supporting existing models which used word-based distributional information, our CAs demonstrated that there is lexical information in the input from which children can learn aspects of the adult-like syntax-relevant semantics necessary to constrain their choice of locative structure.
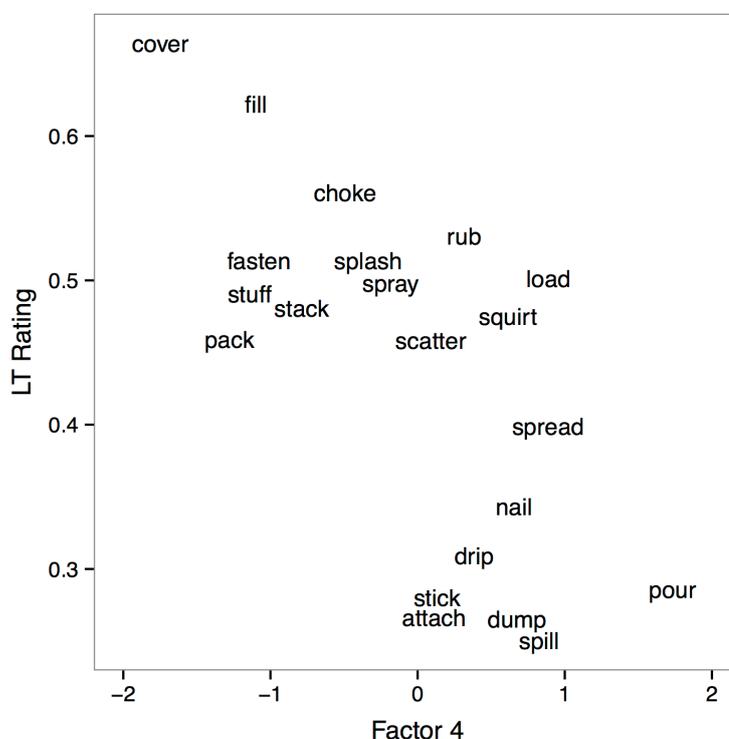

Figure 5. LT bias ratings by Factor 4 in Post-Two Words CA for 21 locative verbs. LT-biased verbs cluster in the top right-hand corner of the figure and TL-biased verbs cluster in the bottom left-hand corner. Two verbs (*attach, spray*) have been jittered for legibility.

Table 2. Correspondence analyses with different input sets. R^2 and related values were obtained from regression against LT Rating using each CA's best factor.

CA name	Target verbs classified	Count types	Count tokens	Linear model with best factor			
				F	df	p	R^2
Full	21	3,277	52,901	5.01	(1,19)	.037	0.21
Post-verbal	21	2,139	21,248	11.59	(1,19)	.003	0.38
Preposition-only	21	7	2,239	10.00	(1,19)	.005	0.35
Two Post-verbal	21	992	8,776	16.65	(1,19)	<.001	0.47
Transitive	21	912	6,771	8.26	(1,19)	.01	0.30
Proportion LT	21		1,680	5.99	(1,19)	.024	0.24

The Full and Post-verbal CA explained less variance than the 47% in the Two Post-verbal Words CA, accounting for 21% and 38% respectively. This is likely because the Full CA and Post-verbal CA treated all post-verbal words as equivalent, leading to a loss of sentence position information. In contrast, the Two Post-verbal Words CA only received the two adjacent, post-verbal words, which made these words position-specific. Because prepositions are useful cues for verb class identification, the Preposition Only CA accounted for a significant amount of variance (35%). Nonetheless, this was 12% less than the Two Post-verbal Words CA, in part because the Preposition Only CA used only 2,239 word tokens, while the Two Post-verbal Words CA used 8,776 due to the frequent transitives in the input. Even without the clear cues provided by the prepositions, the Transitive analysis explained 30% of the variance in the ratings, demonstrating that the distribution of post-verbal articles and nouns was a reliable cue to verb class. Finally, to investigate whether a learner could use syntactic bootstrapping to identify a verb's semantic class, the Proportion LT analysis used the proportion of LT structures in the hand-coded corpus to predict LT ratings. This analysis explained only 24% of the variance – about half of the variance explained by the best Two Post-verbal Words CA. The weakness of this result could be due to in part to the smaller input set (1,680 utterances, around a fifth of the size of the input to the Two Post-verbal Words CA). This sparse input simulated the fact that young children may not correctly parse every single utterance that they hear, just as adults also do not always correctly parse utterances (Ferreira, 2003). In addition, our adult input was TL-biased, which may have weakened its predictive power with respect to learning LT-biased verbs. The TL bias in particular poses a problem for accounts like syntactic bootstrapping, which must explain how children learn LT-biased verb meaning from structurally-inappropriate input. While the Proportion LT and Full CA models yielded only medium effect sizes (Cohen, 1992), the effects sizes for the other lexical CAs were large. This suggests that word-based distributional learning would be a more efficient than a structure-based approach to learning verb classes that predict structural choices.

The best, Two Post-verbal Words CA explained nearly half of the variance associated with 21 verbs in adult ratings. This is a striking result given that adults' ratings were based on years of experience with locative verbs' syntax and semantics, whereas this CA used a much smaller, heterogeneous child-directed corpus that did not filter out non-verb uses. While prepositions were expected to predict locative structures, we found that post-verbal words without prepositions also predicted 30% of the variance associated with verb class ratings. This usefulness of transitive input in locative verb bias acquisition poses a challenge for one particular version of syntactic bootstrapping, which argues that children use the number of arguments, rather than a formal syntactic structure, to bootstrap the meaning of verbs (Fisher, Gertner, Scott & Yuan, 2010). This argument-counting mechanism would only use three-argument input when learning about locative verbs. Since transitive structures have one fewer argument than locative structures, this mechanism would be unable to take advantage of locative verbs in transitive utterances. For theories of syntactic bootstrapping to assign locative verb meanings from transitive structures (e.g., assign different verb classes for *the boy is gorping the blanket* and *the boy is pilking the bed*), they would need to incorporate a lexical distributional mechanism like the CA that we have used here, because LOCATION and THEME transitives have the same structure and number of nouns. Since many structures that are used in syntactic bootstrapping studies also differ in words (e.g., *the boy is gorping **the** girl, the boy **and** girl **are** gorping, the boy is gorping **with** the girl, **he** is gorping **her***), this lexical distributional account of verb class assignment could explain a wide range of syntactic bootstrapping results without using structure-based heuristics.

3.1 A Connectionist Model of Acquisition of the English Locative Alternation

Our corpus analysis documented how locative verbs and structures are used over development, and the correspondence analyses showed that there is sufficient information in the input for children to acquire verb classes from word co-occurrence regularities. However, there is more to verb class acquisition than simply acquiring verb classes: children must simultaneously learn structures and map them to meanings in order to generate grammatical sentences. In the next section we examine computational models of syntax acquisition to better understand how these multiple representations might be learned and integrated.

Existing computational models of verb-structure acquisition assume that children have a limited set of verb class hypotheses and use adult-like syntactic structures and rich semantic features to learn verb classes. For example, Niyogi (2002) developed a Bayesian model of locative verb class acquisition which assigned verbs to a small set of innate verb classes based on pre-existing syntax-relevant semantic features (e.g., *container is full*) or structural information (e.g., X *verbed* Y with Z). In contrast, Perfors, Tenenbaum and Wonnacott's (2010) Bayesian model did not assume innate verb classes. Instead, it specified hyper-parameters that constrained the types of hypotheses that were entertained and used adult-like syntactic structures and verb class semantic features to set these parameters. Alishahi and Stevenson's (2008) Bayesian model of verb argument structure acquisition extended these assumptions by enriching their syntactic structures with a large number of fine-grained semantic features (e.g., *change of state*) for learning. Because there is no automated way to label corpora for syntactic structure or fine-grained semantics, these models all took corpora that had been manually labelled with structural or semantic features by adult linguists. Children, however, learn language from utterances without syntactic category labels, and situational semantic features may not be available for low-frequency verbs (e.g., *litter*; see Section 1). Based on these models, it is not clear how verb-structure regularities are learned if children do not have input that approximates these adult syntactic/semantic codings. Furthermore, no existing model has predicted the main finding in our corpus work: that children initially

have a strong bias towards the TL structure which gradually diminishes as they acquire verb classes.

To develop an explicit model of locative acquisition, we adapted a connectionist model of sentence production and acquisition called the Dual-path model (Chang, 2002). This model lends itself well to the current work because it has been shown to acquire the locative alternation and can explain structural priming with this alternation (Chang, Bock & Goldberg, 2003; Chang, Dell & Bock, 2006). In addition, the model can explain syntactic bootstrapping findings in preferential looking studies (e.g., Naigles, 1990). For example early in development, the model was able to map causative meanings to transitive sentences and non-causative meanings to *with*-intransitives (Chang et al., 2006). Furthermore, its predictions about early structural biases in preferential looking, namely that causative-transitive mappings are learned earlier than non-causative-intransitive mappings, have been confirmed in several syntactic bootstrapping studies in different languages (Gertner & Fisher, 2012; Matsuo, Kita, Shinya, Wood, & Naigles, 2012; Noble, Rowland, & Pine, 2011). More broadly, the model provides a general account of language acquisition and sentence production, learning typologically-different languages such as English, Japanese, and German (Chang, 2009; Chang, Baumann, Pappert, & Fitz, 2014) and successfully models a range of findings from the child and adult production literature, for example structural priming, conceptual/lexical accessibility, heavy NP shift and the accessibility hierarchy (Chang, 2009; Chang et al., 2006; Fitz, Chang, & Christiansen, 2011; Rowland, Chang, Ambridge, Pine, & Lieven, 2012).

In our simulations, the Dual-path model learned syntactic representations from input pairs consisting of a message (which simulated the meaning that is being conveyed) and the heard sentence.

An example of a LT locative message-sentence pair is shown in (3).

3) Sentence: *the man sprayed the wall with water*
    Message: ACTION=SPRAY, AGENT=MAN, THEME=WATER, LOCATION=WALL

In contrast, a TL locative message-sentence pair might be represented as in (4):

4) Sentence: *the man sprayed water onto the wall*
    Message: ACTION=SPRAY, AGENT=MAN, THEME=WATER, LOCATION=WALL

The message contained role-concept links that encoded message-specific mappings between the thematic roles (e.g., ACTION, AGENT, LOCATION, THEME) and concepts (e.g., SPRAY, MAN, WALL, WATER) in an event. Because the same message can correspond to both the LT and TL structures (3,4), thematic roles do not determine grammatical function. For example, the syntactic object is the LOCATION *wall* in the LT locative (3) and the THEME *water* in the TL locative (4). Likewise, the oblique object (the noun in the prepositional phrase, e.g., *the man sprays the wall with **water***) could be either the LOCATION or the THEME dependent on structure. In addition to the role information, the message also contained event-semantic information that signalled the number of roles involved in an event. For example, if the event-semantics signalled an AGENT, LOCATION and THEME role, then the model

learned that either a LT or TL structure was possible. Critically, the message information did not provide the fine-grained verb class semantics that is sometimes assumed to support the acquisition of verb-specific biases for locative structures (e.g., change of state; Ambridge et al., 2012; Levin, 1993; Pinker, 1989).

To test whether distributional regularities in the model's input could support locative verb class acquisition, we created a grammar in which groups of verbs occurred systematically in LOCATION- or THEME-biased structures. Verbs were grouped into five classes. Nonalternating class A verbs occurred in the LT structure only and nonalternating class E verbs appeared in the TL structure only. The other three classes B, C and D alternated between the L and T transitive structures in intermediate degrees (75%, 50%, and 25% respectively). Since these three classes were never experienced in a full locative, the model had to learn how these verbs were biased in the locative from the distributional word regularities in transitives. To instantiate the idea that verb biases must be learnable without clearly biased situational semantics, the thematic roles AGENT/LOCATION/THEME for utterances with class B, C and D verbs were left empty in training. Thus the structural knowledge of the locative could only be learned from class A and E verbs, and verb-structure biases for B, C and D could only be acquired from distributional analysis of transitives (as in the correspondence analyses described in Section 2.2).

3.1.1 Model Architecture

The Dual-path architecture had two separate processing pathways: the sequencing system and the meaning system (bottom and top parts of Figure 6, respectively). The sequencing system learned syntactic categories and ordering constraints among those categories. The meaning system activated thematic roles which in turn triggered concepts specific to particular messages. The sequencing system is depicted in the bottom part of Figure 6 and consisted of a Simple Recurrent Network (SRN; Elman, 1993). The SRN had a feed-forward architecture that mapped from the Previous Word layer via a Hidden layer to the Produced Word layer. The model used the previous word to activate Hidden unit representations which helped it to predict the next word in a sentence (centre of Sequencing System panel, Figure 6). Between the Hidden and the Previous Word / Produced Word layers were Compress layers: smaller layers that forced the model to cluster words into syntactic categories based on distributional regularities (Elman, 1993). In addition, a Context layer held a copy of the previous Hidden layer representation, allowing the model to encode longer-distance regularities (Elman, 1990; 1993). To learn the language, the SRN generated a prediction for the next word in its Produced Word layer and compared that output with the next word in the "heard" utterance. The mismatch between the two generated error, which was backpropagated through the network to change the weights (Rumelhart, Hinton, & Williams, 1986), so that these weights were better able to support next word prediction. The SRN therefore learned syntactic categories and structures because these representations were useful for next word prediction.

The meaning system is depicted in the top panel

of Figure 6. Unlike the learned weights in the rest of the model, the message was instantiated with fast-changing weights between the Role layer and the Concept layer (bottom right of Meaning System panel, Figure 6). The meaning system assumed that when a speaker generates a message (e.g., *the man sprays the wall with water*), the Role-Concept links are set before sentence production begins in order to encode the intended message (e.g., ACTION=SPRAY, AGENT=MAN, THEME=WATER, LOCATION=WALL). The Concept layer was connected to the Produced Word layer by weights which the model learned gradually through cross-
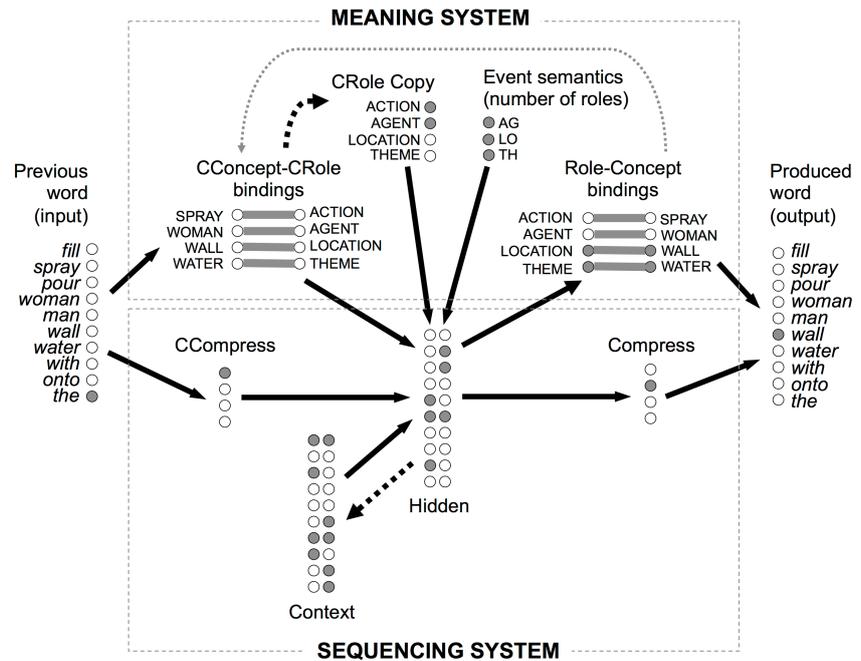
situational learning (Smith & Yu, 2008). The Hidden layer was connected to the Role layer to enable the model to learn to activate roles at particular positions in sentences. The Role-Concept message allowed the model to generalise its

Figure 6. Architecture of the Dual-path model at the post-verbal position (e.g., *the man spray the…*). The Role-Concept bindings activate the LOCATION-WALL and THEME-WATER mappings, and via a winner-takes-all mechanism, the word *wall* is selected at the post-verbal position, which eventually leads to production of the LT locative structure. Learned links are indicated by solid arrows, copy links are indicated by dashed arrows, and links set by the message at the beginning of sentence production are indicated by bold lines.

knowledge in order to generate novel sentences. For example after learning to produce sentences like the *the man sprays the wall with water* from a message, if the message was changed to ACTION=SPRAY, AGENT=FAIRY, THEME=COFFEE, LOCATION=UNICORN, the same sequence of role activations would produce the novel sentence *the fairy sprays the unicorn with coffee* (see Chang, 2002, for a discussion of how the fast-binding message addresses the issue of compositionality in connectionist models; Fodor & Pylyshyn, 1988; Marcus, 2003).

In structures with several alternative word orders, multiple role units could become activated at the *choice point* where the structures diverge. In the locative, the choice point occurs at the post-verbal noun, where the model chooses to activate the LOCATION or THEME role. For example at the post-verbal position depicted in Figure 6, the model has learned that *spray* can be followed by either a LOCATION or a THEME, and consequently activates both roles after the verb. This leads to competition between the concepts which fill those roles (e.g., WALL and WATER), which spreads to the corresponding word units (e.g., *wall* and *water*). Because the word layer used a winner-take-all mechanism, a small difference in activation could lead one unit to be selected over another. Thus, this part of the architecture instantiated the competition between referents and learned concept-word associations specified by both theories of word learning (McMurray, Horst, & Samuelson, 2012) and word production (Gordon & Dell, 2003).

Because lexical competition at the choice point can influence structural choice (Bock, 1982; Chang, 2009), the model had to relay information about its lexical

choices back to the sequencing system. It did this by passing back the winning Produced Word output as the next Previous Word input. To influence structural choice, however, the model needed to know the role of this winning word (Chang, 2002). For example in order to produce a full locative after starting with the fragment *the man spray the wall*, the model had to determine that *wall* mapped to LOCATION in the present message in order to activate the word *with* and the THEME (e.g., *with water*). The role of the previous word was provided by the CConcept-CRole bindings (bottom left, Meaning System panel, Figure 6; the "C" specified that this layer operated in the "comprehension" direction). In the above example, when the *wall* unit was activated in the Previous Word layer, the model used learned links to activate the WALL concept in the CConcept layer. Because the wall was a location in this particular message, the LOCATION unit was activated in the CRole layer. This signalled to the sequencing system that the next word should be *with* and that the LT structure should be produced. In addition, a CRole Copy layer averaged activation of its own previous state and the previous CRole state, extending the model's memory for the roles produced earlier in the sentence (top centre-left, Meaning System panel, Figure 6).

Finally, in order to produce a correct sentence the sequencing system needed to know how many arguments the sentence requires. For example, *the man spray the wall* could be a complete transitive sentence, or it could continue (e.g., *the man spray the wall with water*). To provide the model with this information, Event-Semantics units encoded the number of arguments in the message (top centre-right, Meaning System panel,

Figure 6). For example in an LT locative sentence, the Event Semantics units for the AGENT (AG), THEME (TH), and LOCATION (LO) were activated. Details of model parameters and inputs are provided in Appendix B.

In summary, the model learned its representations through error-driven learning based on its next word predictions using weights from both meaning and sequencing systems. We think that the model's sequencing system developed abstract syntactic representations, because the model can account for a wide range of structural priming effects (Chang et al. 2006), in particular priming effects between structures that differ in words and thematic roles (Bock & Loebell, 1990; Bock 1989). The same message-based prediction system supported production of the model's own messages (Dell & Chang, 2014). The event-semantics signalled which alternative structures were available and the "comprehension" message allowed the model to alternate between structures based on the words that were produced at choice points (e.g., the post-verbal position in locatives).

3.1.2 Comparison of models of locative acquisition

The goal of our simulations was to understand how interactions between the input and the architecture of a language learner could explain three patterns that we found in our corpus analysis. First, we found an early bias towards the TL locative structure. Second, the TL bias diminished over development as verb classes were acquired. Third, the larger and more balanced nature of transitive uses in the input suggested that these structures were an important source of information about locative verb biases. Our main simulation, the Transitive Generalisation model, captured these three patterns. To understand which features of the input or architecture contributed to the success of this model, we developed several variants. The No Post-verbal Theme model showed that the TL bias early in development was due to frequent input structures with post-verbal THEMES (e.g., transitives). Next, the No-Article Omission model demonstrated that verb class acquisition depended on utterances which omitted mass noun articles (e.g., *spray water on the wall*). Finally, the No CCompress model highlighted the role of the model's architecture in creating verb classes. Together, these simulations showed that the patterns in our corpus analysis could be explained by the interaction of the model's architecture and features of the input.

Table 3. Percentage of LT and L structures in the training set for five verb classes.

	Class A	Class B	Class C	Class D	Class E
Locative	100%	-	-	-	0%
Transitive	-	75%	50%	25%	-

The input to the Transitive Generalisation model formed the basis for all the models that will be compared in this section. We generated the input using a training grammar with five locative verb classes (A, B, C, D, E; Table 3) that appeared in LOCATION or THEME structures (i.e. LT locative, TL locative, L transitive, or T transitive) with varying frequency. Class A verbs only occurred in LT locative structures; for example the class A verb *fill*

appeared in sentences such as *the woman fill the cup with water*. Class E verbs only occurred in TL locative structures; for example the class E verb *pour* appeared in sentences such as *the woman pour water into the cup*. Classes B, C and D simulated alternating locative verbs. Our corpus data suggested that children may not encounter all locative verbs in locative structures, and may therefore have to learn verbs' biases from their more frequent transitive uses. For example, in our corpus data the verb *shake* never appeared in a locative structure, making it impossible to learn its bias from locatives. However, *shake* did appear in transitive structures: 15 times in the T structure and three times in the L structure. The transitive distribution of *shake* (e.g., *you just shake your head*) suggested that it is a TL-biased alternating verb (class D). To examine whether the Dual-path model can use the distribution of locative verbs in transitive structures to infer verb class, verbs in classes B, C and D only occurred in transitive structures: class B occurred in the L structure 75% of the time (and the T structure 25% of the time), class C 50% of the time, and class D 25% of the time. To ensure that the model learned to assign verbs to verb classes based on distributional input alone, we removed all AGENT, LOCATION and THEME thematic role and concept information for these three alternating locative verb classes.

The model's small vocabulary meant that it could learn word-to-word constraints that could support structural choice at test. For example, encountering *the man dripped water* in training increased the likelihood of the model producing, say, *the boy dripped water onto the table* at test. This point is critical: learning word-to-word regularities would allow the model to produce the appropriate locative structure for each individual verb without needing to learn verb classes. To ensure that the model's performance was not purely lexical in nature, transitive training and locative test items included different nouns for each verb class so that at test, the noun/verb combinations the model encountered for a given alternating class were completely novel. Thus, classes B, C, and D occurred with half of the eight LOCATION nouns in training and the other half at test, and likewise for the THEME nouns (note, each verb class contained different combinations of nouns). In verb classes A and E, all eight LOCATION and eight THEME nouns occurred together in training, enhancing generalisation in classes B, C and D. Thus, rather than learning lexical constraints like "*dripped* is followed by *water*", the model learned that *dripped* was frequently followed by THEME nouns.

The model's input language included several features which made it closer to English language input. First, because mass noun THEMES can appear without articles in English (e.g., *the man sprayed water onto the wall*), we deleted 25% of THEME noun determiners in the input. Since 75% of the time these nouns occurred with determiners, determiner production was still the default option. A similar rate of determiner omission (22%) was found in a corpus analysis of all the tokens of 10 CDI mass nouns (*bread, cake, cheese, coffee, food, juice, milk, pasta, toast, water*; Hamilton, Plunkett & Schafer, 2001) in the CHILDES Manchester corpus (Theakston et al., 2001). Second, half the LOCATION nouns in TL sentences

were treated as containers and occurred with the preposition *into*, while the other half were treated as surfaces and occurred with the preposition *onto*. Third, in addition to class A and E locatives and B, C and D transitives, the training grammar also included intransitive, transitive, double object dative, and prepositional dative structures. None of these structures contained verbs from classes A – E. We labelled these transitive structures *non-locative transitives* to distinguish them from the "locative" transitives which included verbs from classes B, C, and D. For simplicity, intransitives, double object datives, and prepositional datives were equally frequent as either class A or class E locatives. Since locative verbs occurred as transitives more frequently than locatives in our corpus, class B, C and D verbs were four times as frequent as class A locatives. The non-locative transitives were 12 times as frequent as class A locatives given that they are the most frequent structure in corpus studies (Roland, Dick & Elman, 2007). All sentences in training and test grammars only used the article *the* and contained no tense or aspect variation.

The test grammar included verb classes A – E, but only included locative message-sentence pairs. Each verb was paired with both LT and TL structures, and the model's accuracy in producing these structures was used to calculate the proportion of LT structures out of all correct utterances. This measure was used as an index of verbs' structural biases. For example, if the model learned that *fill* is LT-biased, then it should correctly produce LT sentences such *the man fill the bucket with water*, but struggle with TL sentences such as *the man fill water into the bucket*. Thirty training sets were created to simulate the variability in the input across different participants, resulting in 30 model subjects. Each training set consisted of 40,000 randomly-generated message-sentence pairs. Test sets consisted of 1,000 randomly-generated message-sentence pairs. The model was trained for a total of 40,000 epochs (weights updates after each sentence) and tested every 4,000 epochs. As children do not always hear sentences with a matching situation, a randomly-selected 50% of the training sentences appeared without a message.

To obtain a measure of overall structural bias for each verb class we compared the sentence the model produced to the target in the test set. Sentences were scored correct if the produced and target sentence were the same word sequence. We calculated the proportion LT out of all LT and TL structures for each verb class at each testing point in each model. This proportion was transformed using the empirical logit transform (Agresti, 2002), which then served as the dependent measure for a mixed effect model with verb class (numerically coded, B = 1, C = 0, D = -1) crossed with epoch (divided by 4000) as predictors, model subject as a random variable and by-subject slopes for verb class and epoch main effect (the maximal model for our dataset; Barr,

Levy, Scheepers & Tily, 2013). All simulations use the same grammar, model parameters, and mixed model analysis, except when specified.

3.2.1 Transitive Generalisation Results

Figure 7 shows that the model correctly distinguished classes B, C and D in the predicted direction (B > C > D; main effect of verb class, beta = 0.12, $\chi^2(1)$ = 42.79, p < .001) and these verb classes separated over development (significant interaction between epoch and verb class; beta = 0.010, $\chi^2(1)$ = 13.64, p < .001). The model exhibited the early TL bias that we found in our corpus analysis (Figure 7; negative logit intercept coefficient beta = -0.74). There was also a significant effect of epoch (beta = 0.030, $\chi^2(1)$ = 14.06, p < .001), which indicated that, like children, the model recovered from this early bias, producing more LT locatives over development. However, early in learning at epoch 8,000, 87% of all correct utterances were TL locatives. Importantly, the model linked the TL bias in our corpus study with the TL overgeneralisation errors in Bowerman's (1982) data. Class E verbs only appeared in the TL structure in training, and the model produced class E verbs exclusively in TL structures at epoch 8,000 (proportion LT = 0). However, although class A verbs always appeared in the LT structure in the input, the model overgeneralised these verbs 10% of the time at epoch 8000 to the TL structure (e.g, *the man fill the chair into the house*; *the father fill the apple onto the apartment*). These overgeneralisation errors mirrored those in Bowerman's (1982) data, and showed that children's overgeneralisation errors could have arisen from the early TL bias. The model therefore provided an explicit account of how a general bias towards one structure can lead to overgeneralisation of that structure, and how those errors decreased as the TL bias reduced alongside the acquisition of verb classes.

3.2.2 The *No Post-verbal Theme* simulation

The TL bias in the Transitive Generalisation model could not have been due to the thematic roles in the locative input as LT and TL structures with thematic roles occurred equally often in verb classes A and E. Instead,
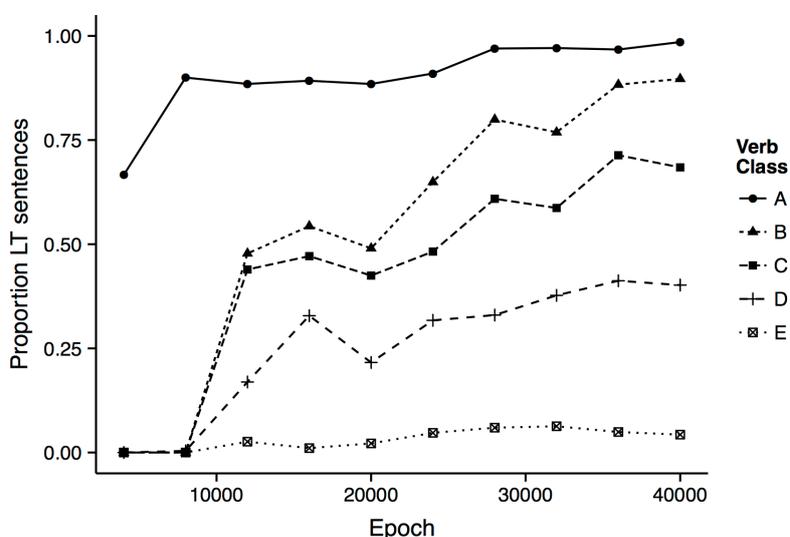

Figure 7. Results of Transitive Generalisation simulation: proportion LT locatives by epoch and verb class
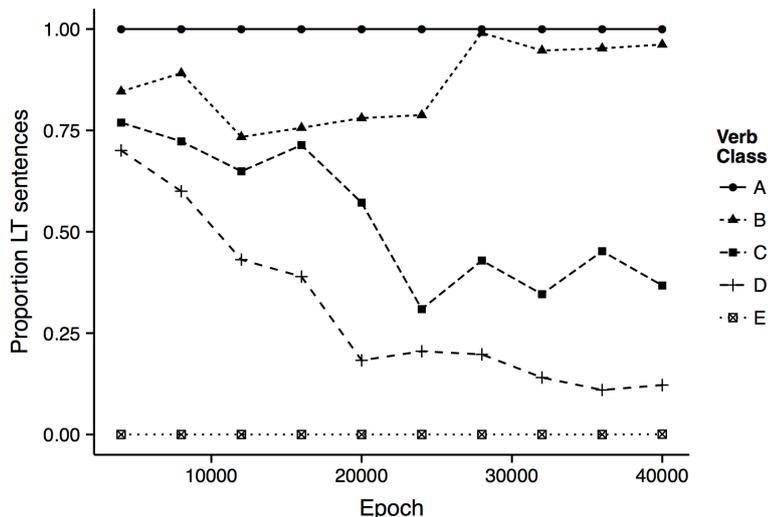
Figure 8. Results of No Post-verbal Theme simulation: proportion LT locatives by epoch and verb class

the post-verbal THEME roles in the frequent non-locative transitives and prepositional datives may have biased the model towards post-verbal themes. To test this possibility, we designed a *No Post-verbal Theme* grammar, which was identical to the Transitive Generalisation grammar (Section 3.2.1) except that we removed all non-locative transitives and prepositional datives. Since this model had B, C and D locative transitives in its input, this simulation differentiated verb classes (Figure 8; main effect of verb class: beta = 0.059, $\chi^2(1)$ = 57.66, $p < .001$), and this differentiation increased over development (verb class by epoch interaction: beta = 0.024, $\chi^2(1)$ = 52.24, $p < .001$). Unlike the Transitive Generalisation simulation, however, this simulation did not exhibit an early TL bias (intercept greater than 0.5; logit beta = 0.49) and therefore did not overgeneralise LT-biased class A verbs into the TL structure. This demonstrated that the TL bias in the Transitive Generalisation model was due to in part to thematic roles in frequent non-locative transitive and prepositional dative structures, which created an early bias towards post-verbal THEMES.
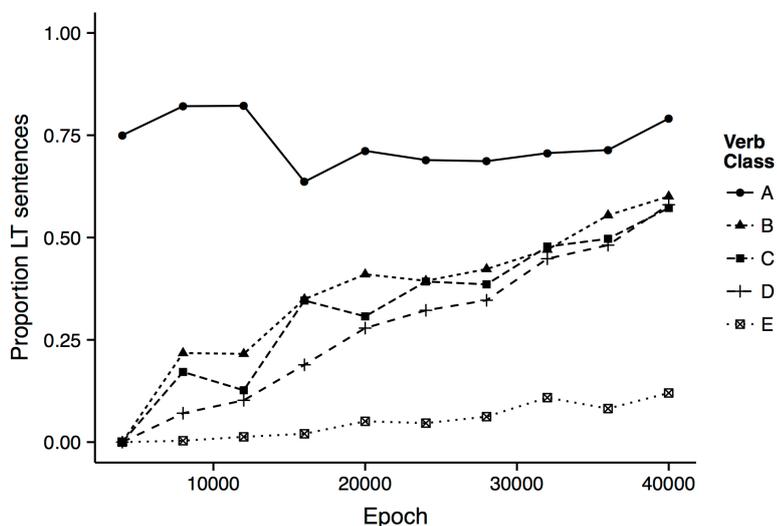

Figure 9. Results of No Article Omission simulation: proportion LT locatives by epoch and verb class

3.2.3 The *No Article Omission* simulation

In developing the model, we discovered that it would not learn verb classes without the article omission often found on THEMES in the locative (e.g., *the man sprayed water on the wall*). To simulate the omission found in English, the Transitive Generalisation model omitted articles 25% of the time for mass nouns in the THEME role. To test whether this was important for verb class acquisition, we created a *No Article Omission* model in which all nouns were preceded by the article *the*.

As depicted in Figure 9, because this grammar contained non-locative transitives and datives, there was an early TL bias (intercept beta = -0.92) that diminished over time (main effect of epoch: beta = 0.027, $\chi^2(1)$ = 20.09, $p < .001$). However, although the model differentiated verb classes overall (main effect of verb class: beta = 0.16, $\chi^2(1)$ = 10.42, $p = .0012$), it did not do so in a way that corresponded to the input distribution. For example, class A was 100% LT in the input, but the model produced around 75% LT in its output at the end of training. The difficulty in linking verbs and post-verbal nouns across the article relates to a general difficulty that SRNs have in learning non-adjacent regularities (Bengio, Simard, & Frasconi, 1994). In the *No Article Omission* model, all verbs occurred with the article *the*. The SRN therefore learned a single class in the Hidden layer for all locative verbs, because they all predicted the same next word *the*. On the next timestep, after the Hidden layer activation was copied back to the Context layer, the model tried to use the Context layer's activation to predict the LOCATION or THEME noun. Since this Context activation only represented a single verb class, it was difficult for the model to learn different verb-specific biases. In the Transitive Generalisation model, LT/TL-biased verbs differed because TL structures were more likely to omit post-verbal articles. The SRN learned that class B verbs tended to be followed by *the*, while class D verbs were occasionally followed by a noun (when the article was omitted). Having learned to distinguish B and D verbs in its Hidden layer, this information would also be copied back to the Context layer, allowing the model to exhibit verb bias even when there was an article intervening, as in the test items used here.

3.2.4 The *No CCompress* simulation

The previous simulations suggested that input shaped the model's verb class acquisition. However, the model's architecture also influenced the categories that it learned. To examine the influence of the architecture, we
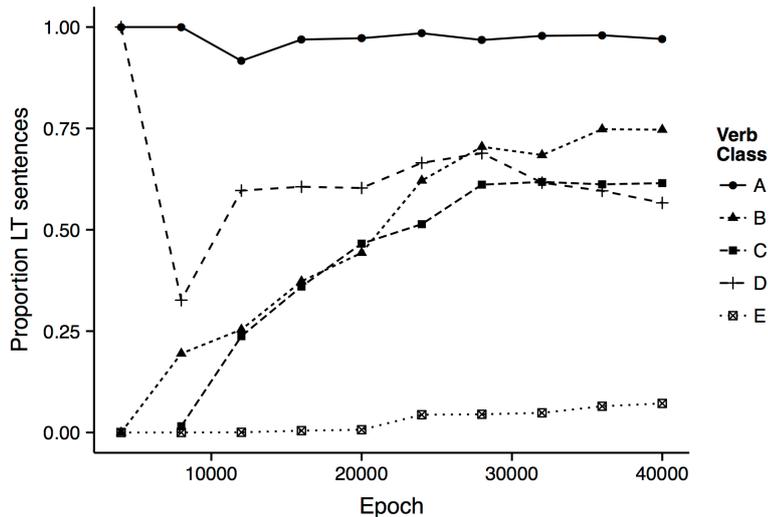
Figure 10. Results of No CCompress simulation: proportion LT locatives produced by epoch and verb class

removed the most probable locus of verb classes in the model, the CCompress layer between the Previous Word layer and the Hidden layer (Figure 6, bottom panel, left), and replaced it with direct connections between the Previous Word layer and the Hidden layer. As depicted in Figure 10, while the *No-CCompress* model differentiated verb classes over development (verb class by epoch interaction: beta = 0.020, $\chi^2(1) = 35.82$, $p < .001$), the final bias for each class did not reflect the separation of biases in the input. Again, the post-verbal THEMES in the input generated an early TL bias that dissipated over time (main effect of epoch: beta = 0.017, $\chi^2(1) = 9.79$, $p = .0018$). This simulation demonstrated that the input alone was not sufficient to explain verb class development without some architectural constraint that motivated the compression of verbs into classes.

3.3 Grammaticality Ratings
A limitation of our corpus study is that we used children's increasing LT production as an indirect indication that verb classes have been acquired. Although the corpus data were too sparse to show the separation of verb
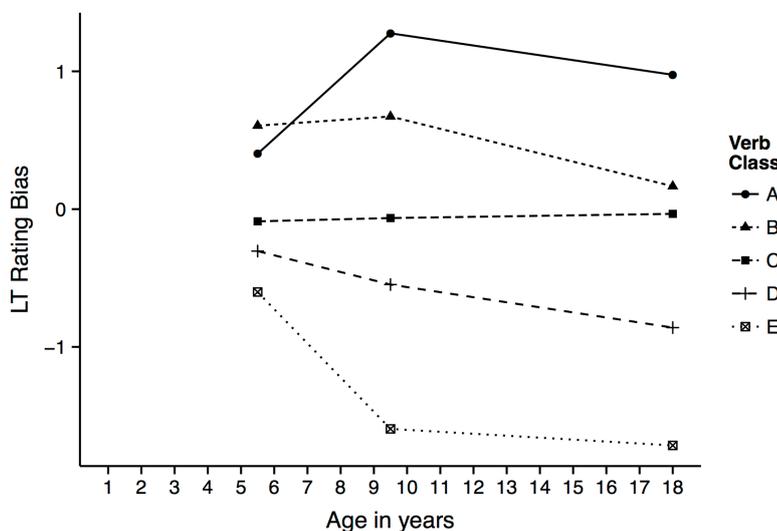
classes over development, experimental data from grammaticality ratings studies exhibited gradual differentiation of locative verb classes over development (Ambridge et al., 2012; Bidgood et al., 2014). In these studies, participants were asked to rate the grammaticality of locative verbs in LT or TL structures on a Likert-type scale (see Ambridge, 2011). We reanalysed these data using a similar method to the model analyses. We created an effect size measure of the verb's preference for LT (*LT rating bias*) by computing Cohen's *d* from the difference in the rating of the verb in LT and TL structures divided by the standard deviation for that verb.

Figure 11 depicts LT rating bias for adult ratings for five of Pinker's (1989) semantic verb classes, selected such that they varied from strongly LT-biased to strongly TL-biased. These classes were labelled A, B, C, D, and E based on their LT bias according to adult ratings (e.g., class A was strongly LT-biased). To examine whether this separation grew over development, we coded verb class numerically (A = 2, B = 1, C = 0, D = -1, E = -2) and fitted a mixed effects model to LT rating bias with verb class and age crossed, verb as a random variable and by-verb slopes for age. A main effect of verb class on LT rating bias demonstrated that ratings differed for verb classes (beta = 0.204, $\chi^2(1) = 64.82$, $p < .001$), and an age by verb class interaction confirmed that differences between the classes grew over development (beta = 0.026, $\chi^2(1) = 20.95$, $p < .001$).

To simulate these grammaticality ratings in the Dual-path model, we used sum of squared error between target test sentences and produced sentences in the Transitive Generalisation simulation. This measure has previously been used as a proxy for graded measures of grammaticality (Chang et al., 2006; MacDonald & Christiansen, 2002). Mean difference in error for each verb class in LT and TL structures is depicted in Figure 12, with positive scores indicating a bias towards LT (i.e., larger error in TL). We submitted error difference to a mixed effects model with verb class (A = 2, B = 1, C = 0, D = -1, E = -2) and epoch crossed, model subject as a random variable and by-subject slopes for verb class crossed with epoch. A significant verb class by epoch interaction (beta = 0.0099, $\chi^2(1) = 512.58$, $p < .001$) showed that verb classes separated across development. The main effects of verb class and epoch were also significant (respectively, beta = 0.013, $\chi^2(1) = 84.62$, $p < .001$ and beta = 0.011, $\chi^2(1) = 8.18$, $p = .0042$). Data from the model therefore reflected the data from the grammaticality ratings studies, with verb classes exhibiting graded grammaticality in LT and TL constructions.
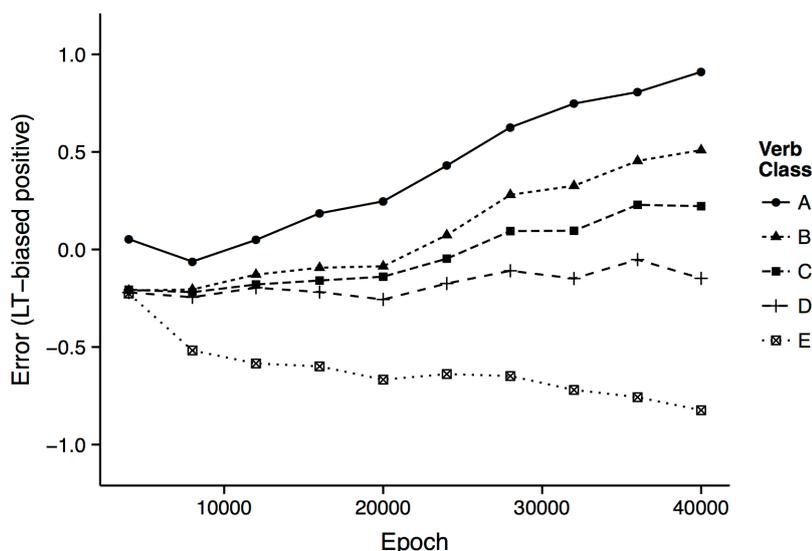

Figure 11. LT rating bias from participants aged 5 to 18 for five verb classes.

Figure 12. Mean error difference in Dual-path model by epoch and verb class. Positive difference indicates bias towards LT (larger error for TL).

Taken together, these simulations captured several important features of our behavioural data and suggested how the input may be involved in creating these patterns. The model's input incorporated a training bias: locative verbs were biased to appear with either LOCATION nouns (e.g., class B) or THEME nouns (e.g., class D) in transitive structures. The model first acquired syntactic structures and gradually used distributional learning to acquire verb classes. Because syntactic structures were learned first, the model exhibited a TL bias for locative structures based on the locative's shared argument preferences with non-locative transitives and prepositional datives. Verb classes emerged gradually over development, mirroring changes in class-based grammaticality ratings in judgment studies (Ambridge et al., 2012). Critically, the model demonstrates that these locative verb classes can be learned from transitive structures, which are more frequent and balanced in our corpus analysis and which the CA comparisons demonstrate are a significant predictor of adult verb classes. Finally, these results were generated by a general-purpose model that can explain a range of language acquisition and production phenomena.

In contrast to existing models of verb class acquisition (Alishahi & Stevenson, 2008; Niyogi, 2002; Perfors, Tenenbaum & Wonnacott, 2010) which assumed that children use adult-like semantics and syntactic features to acquire verb classes, the present work suggested that children can assign some verbs to classes based on word co-occurrence information. However, because language acquisition in the present model included a message with various kinds of meaning, it is important to characterise how our account differs from these accounts. All theories and models agree that a locative structure will only be produced at test when a speaker has a message for describing a three-participant locative scene; for example, speakers must have semantics for selecting a locative verb. In addition, speakers must identify an AGENT, a LOCATION, and a THEME. If this information is unbiased, as in our

computational simulations, both the LT and TL structures are licensed. All of these models, including the Dual-path model, can simulate how children use situational and syntactic information to acquire language representations. However, the Dual-path model adds an additional mechanism: a word-based distributional learning ability. This allows it to explain how young children – who are not fully adult-like in the ability to parse syntax – classify verbs that do not occur in locatives in proportion to their bias or in situations with clear semantic constraints.

4.1 General Discussion

Verb-structure overgeneralisation errors provide an important window into syntactic representations in development. Children do not encounter overgeneralisations in their input but nonetheless produce them, so such errors cannot be trivially explained as the result of learning. However, because children's overgeneralisations are rare, it has been difficult to trace the mechanisms that underlie them. In this work, we tried to better understand these errors by looking at the structural biases in normal utterances. We examined the use of a wider range of verbs in child language corpora than in previous studies (e.g., Bowerman, 1982; Iwata, 2008; Laffut & Davidse, 2002). We discovered an early TL bias which gradually abated over development, and which is also seen in experimental studies of the locative as well as overgeneralisation errors in corpora (Bowerman, 1982; Gropen et al., 1991a, 1991b). Our modelling work showed that this TL bias created overgeneralisation errors with LT-biased verbs. Together, these findings suggested that these biases in correct utterances could help explain overgeneralisation errors early in development.

In our corpus analysis, we found that some verbs do occur in the transitive but not the locative, and that the structural biases of locative verbs in the transitive structure were more consistent over development than the biases of locative verbs in locative structures. Our correspondence analyses demonstrated that using just two post-verbal words without semantic or syntactic labelling, we could cluster verbs in a way that strongly predicted their structural biases as measured by adult ratings. These findings suggested that transitive input provides an important source of information about locative verb classes. The Dual-path model then provided computational evidence that this input can indeed be used to learn verb classes. These two models are very different mechanisms, but they share the fact that their representations for verbs are shaped by the lexical distribution, that is, all the words that co-occur with the verb (Redington et al., 1998). This distinguishes a distributional approach from other lexical learners that use more non-distributional cues (e.g., frequent frames, Mintz, 2003). The distributional approach allows these

models to take advantage of the large amount of input in transitive utterances for learning locative verbs.

Various mechanistic accounts of how verb classes might be learned from semantic and syntactic information have been proposed (Alishahi & Stevenson, 2008; Niyogi, 2002; Perfors, Tenenbaum & Wonnacott, 2010). Learning in these models was quick, because their cues for verb bias were adult-like and optimal for solving this problem. In contrast, the Dual-path model learned more slowly, as it had to develop both syntactic structures and verb classes from word sequences and then gradually learn the link between them. This created developmental preferences like the TL bias which was also present in our corpus analysis. The model's message functioned like Pinker's (1989) broad range rule, allowing verbs that had only appeared in the LT structure to generalise to the TL structure, creating early TL overgeneralisations. Later in development, the model differentiated the verb category into a set of verb classes that differed in the way in which they predicted post-verbal words (as in the CAs). Once it had formed these classes (akin to Pinker's narrow range rules; Pinker, 1989), the model used them to accurately produce LT structures. As LT production increased over development, the TL bias reduced, leading to the model's retreat from overgeneralisation. Furthermore, given that the same model has successfully captured syntactic bootstrapping behaviour in preferential looking studies (Chang et al., 2006), this work suggests that the Dual-path model may provide a general account of a range of distributional learning phenomena in acquisition.

Beyond offering a unified account of verb class acquisition, the model also suggested how some features of linguistic typology might be related to the limitations of sequence learning mechanisms. We found that verb-structure links were difficult for the SRN to learn when the verb and post-verbal noun were always separated by an article. When a more natural, English-like grammar was used in which mass noun THEMES optionally omitted articles, the model consistently learned verbs' structural biases. Other verb bias alternations such as the noun phrase/sentence complement alternation (e.g., *the man wrote* [*that*] *the story* [*was interesting*]; Garnsey, Pearlmutter, Myers & Lotocky, 1997) also involve variability in the post-verbal position (e.g., optional *that*), suggesting that such variation may play a role in verb class acquisition more generally. Humans also find it difficult to learn non-adjacent dependencies in artificial grammar learning tasks and variation can modulate this ability (Gómez, 2002; Gómez & Maye, 2005; Newport & Aslin, 2004). This is particularly relevant given that SRN models provide some of the best accounts of learning in these tasks (Chang, Janciauskas & Fitz, 2012). Finally, when the CCompress layer was removed from the SRN, the model was unable to learn human-like verb class behaviour, suggesting that humans may have a language sequence learning architecture that includes a similar bottleneck to force verbs into a small number of classes.

In this work we demonstrate that locative verb class acquisition need not depend on situational information contained in the message. The message is nonetheless a fundamental component of the Dual-path model's broader learning mechanism, with assumptions which have implications for our understanding of language acquisition. It is therefore important to make these assumptions clear. First, the message in production and acquisition modelled subtly different aspects of production and acquisition. In production, the message represents the idea that the speaker wanted to convey. In acquisition, learning involves message-sentence pairs and the message is related to the interlocutor's intended idea (not the speaker's idea). The model assumed that the ability to map messages in production must be learned from message-sentence pairs because without this message constraint there would be no regularity to the mapping between meaning and form (this relates to Fodor's (1975) claim that the combinatorial/productive nature of language depends on a structured language of thought). Thus, the model also assumed that children must have some message structure in order to explain how they ultimately learn structured mappings between meaning and form.

This assumption raises the question of where this message comes from. Our view is that in many situations, children can infer some parts of messages from visual information in scenes. A large body of work indicates that prelinguistic infants and toddlers can identify thematic role related concepts. Causality (Leslie & Keeble, 1987) and intentionality (Behne, Carpenter & Tomasello, 2005) are useful for identifying agents, object motion (Spelke, 1990) is useful for themes, and goal-directed behaviour (Woodward, 1998) signals locations. Even syntactic bootstrapping studies, which are often used to emphasise the difficulty in inferring meaning situationally, provide some evidence for the ability to infer relational meaning such as thematic roles, otherwise children would not be able to look appropriately at a matching causative or non-causative video even for familiar verbs. Clearly children must have some situational understanding of the world in order to develop structured language representations. The critical claim in this paper, however, is that situational and syntactic information are not always available for learning the bias of every single verb that a child knows, and verb class acquisition therefore requires additional mechanisms such as categorisation based on word distribution.

Any distributional account of language has to explain the origin of the distribution or why is the input structured in a particular way. MacDonald (2013) proposed an influential account of how production biases influence distributional regularities which then are learned by language learners (Dell & Chang, 2014). On this view, situational event structure shapes verb semantics in speakers and their semantically-constrained production output creates the distribution. But an account in which verb class is solely dependent on universal semantic constraints in production would predict that cross-linguistic differences would be rare. The cross-linguistic review in Hunter (2008) suggests that these violations of these universal tendencies are not difficult to find, even for verbs which are strongly biased in English. For example in English, *fill* is strongly biased towards the LT structure and is considered to be ungrammatical in TL, but in Dutch, five out of 20 *fill*-class verbs that encode change of state only occur in the TL structure. The model's overgeneralisations of LT-biased class A verbs like *fill* to TL structures early in acquisition

provided a mechanism for explaining the development of this cross-linguistic variability. If there were stable distributional regularities supporting the TL mapping for some LT-biased verbs, then the model would not retreat from these "errors". These generalisations would then form part of the distribution for the next generation of language learners, eventually leading to differentiation between linguistic communities. Thus, we believe that semantics does influence verb-structure regularities, but its effect is transmitted through a distribution that allows violations of universal tendencies.

When parents urge their children to "Do as I say, not as I do", they are encouraging their children to focus on the linguistic stream, rather than their situational behaviour. Here we suggest that children do something like this when they learn locative verb-structure links. We believe that they focus on the linguistic input because words are more abundant than the pairing of those words with appropriate situational meanings. We took this distributional approach in our corpus work, allowing us to use a large corpus of unlabelled, child-directed utterances to classify a wide range of verbs. In addition, our computational model captured changes in linguistic knowledge over development – normal structural biases as well as errors – and showed that these developmental phenomena are in part due to the challenge of abstracting structures and verb classes from simple word sequences. By situating development within an account of language acquisition in which situational input is relatively impoverished and abstraction from linguistic input is slow and gradual, we provide a new perspective on developmental patterns in human language acquisition.

Author Note

References

Abbot-Smith, K., & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science*, *30*(6), 995–1026. doi: 10.1207/s15516709cog0000_61

Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science, 32*(5), 789–834. doi:10.1080/03640210801929287

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

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, *123*(2), 260–279. doi:10.1016/j.cognition.2012.01.002

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. doi:10.1111/j.1551-6709.2009.01055.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8*(6), 492–499. doi:10.1111/j.1467-7687.2005.00440.x

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks 5*(2), 157–166. doi:10.1109/72.279181

Bidgood, A., Ambridge, B., Pine, J. M., & Rowland, C. F. (2014). The retreat from locative overgeneralisation errors: A novel verb grammaticality judgment study. *PLoS ONE*, *9*(5), e97634. doi:10.1371/journal.pone.0097634

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*(1), 1-42.

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition, 31*(2), 163–186.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition,35*(1), 1–39.

Bohannon, J. N., & Marquis, A. L. (1977). Children's control of adult speech. *Child Development*, 1002–1008. doi: 10.2307/1128352

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 319–246). New York: Cambridge University Press.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651. doi:10.1207/s15516709cog2605_3

Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, *61*(3), 374–397. doi:10.1016/j.jml.2009.07.006

Chang. F., Bauman, M., Pappert, S., & Fitz, H. (2014). Do lemmas speak German?: A verb position effect in German structural priming. Manuscript submitted for publication.

Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, *90*(1), 29–49.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. doi:10.1037/0033-295X.113.2.234

Cruttenden, A. (1978). Assimilation in child language and elsewhere. *Journal of Child Language, 5,* 373–378. doi:10.1017/S0305000900007558

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

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120394.

Dumais, S. T., & Landauer, T. K. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211. doi:10.1207/s15516709cog1402_1

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71–99. doi:10.1016/0010-0277(93)90058-4

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164–203.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 143–149. doi:10.1002/wcs.17

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology, 23*(3), 331–392. doi: 10.1016/0010-0285(91)90013-E

Fitz, H., Chang, F., & Christansen, M. H. (2011). A connectionist account of the acquisition and processing of relative clauses. In E. Kidd (Ed.), *The acquisition of relative clauses. Processing, typology and function.* (pp. 39–60). Amsterdam: Benjamins.

Fletcher, P., & Garman, M. (1988). Normal language development and language impairment: Syntax and beyond. *Clinical Linguistics and Phonetics, 2,* 97–114. doi:10.3109/02699208808985246

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Cambridge, MA: Harvard University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*(1), 3–71.

Forrester, M. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood, 9,* 255–278. doi:10.1177/0907568202009003043

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*(1), 58–93. doi:10.1006/jmla.1997.2512

Gathercole, V. (1986). The acquisition of the present perfect: explaining differences in the speech of Scottish and American children. *Journal of Child Language, 13,* 537–560.

doi:10.1017/S0305000900006875

Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition 124*(1). doi:10.1016/j.cognition.2012.03.010

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*(2), 135–176.

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

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*(1), 23–64. doi:10.1207/s15473341lld0101_4

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*(4), 544–569. doi:10.1016/j.jml.2007.01.007

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11*(01), 43–79. doi:10.1017/S0305000900005584

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

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436. doi:10.1111/1467-9280.00476

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206. doi:10.1207/s15327078in0702_4

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science, 27*(1), 1–40. doi: /10.1016/S0364-0213(02)00111-8

Greenacre, M. (2007). Correspondence analysis in practice. Boca Raton: Chapman & Hall. doi:10.1201/9781420011234

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991a). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition, 41*(1), 153–195. doi:10.1016/0010-0277(91)90035-3

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991b). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language, 18*(1), 115–151. doi:10.1017/S0305000900013325

Gruber, J. S. (1965). *Studies in lexical relations.* (Doctoral dissertation). Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/13010

Hamilton, A., Plunkett, K., & Schafer, G., (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: Lower scores in the UK than the USA. *Journal of Child Language, 27,* 689-705.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of*

Memory and Language, 48(2), 281–303. doi: 10.1016/S0749-596X(02)00516-8

Henry, A. (1995). *Belfast English and Standard English: Dialect variation and parameter setting*. New York: Oxford University Press.

Howe, C. (1981). *Acquiring language in a conversational context*. New York: Academic Press.

Hunter, D. C. (2008). Locative alternations: A cross-linguistic look at the syntax-semantics interface. *Unpublished Master's thesis, University of Amsterdam*. Retrieved from http://dare.uva.nl/document/106321

Iwata, S. (2008). *Locative alternation: A lexical-constructional approach*. Amsterdam: John Benjamins.

Jackendoff, R. (1983). *Semantics and cognition* (Vol. 8). Cambridge, MA: MIT Press.

Joanis, E., Stevenson, S., & James, D. (2008). A general feature space for automatic verb classification. *Natural Language Engineering*, *14*(03), 337–367. doi:10.1017/S135132490600444X

Johnson, M. (1986). *A computer-based approach to the analysis of child language data*. Reading, UK: University of Reading.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. doi:10.1016/0010-0285(89)90003-0

Korman, M. (1992). Korman corpus. *CHILDES*: *Child Language Data Exchange System*. Retrieved December 3, 2012, from http://childes.psy.cmu.edu/browser/index.php?url=Eng-UK-MOR/Korman/

Kim, M., Landau, B., & Phillips, C. (1999). Cross-linguistic differences in children's syntax for locative verbs. In *Proceedings of the 23rd Boston University Conference on Language Development* (pp. 337–348). Somerville, Mass.: Cascadilla Press.

Laffut, A., & Davidse, K. (2002). English locative constructions: An exercise in neo-Firthian description and dialogue with other schools. *Functions of Language*, *9*(2), 169–207. doi:10.1075/fol.9.2.04laf

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174. doi:10.2307/2529310

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. doi: 10.1016/S0010-0277(87)80006-9

Levin, B. (1985). Lexical semantics in review: An introduction. In B. Levin (Ed.), *Lexical semantics in review*. *Lexicon Project Working Papers, 1*. Cambridge, MA: MIT Center for Cognitive Science.

Levin, B. (1993). *English Verb Classes and Alternations:*

A Preliminary Investigation. London: University of Chicago Press.

Levin, B., & Rappaport Hovav, M. (1991). Wiping the slate clean: A lexical semantic exploration. *Cognition*, *41*(1), 123–151. doi:10.1016/0010-0277(91)90034-2

Lieven, E., Salomo, D. & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics,* *20*(3), 481–508. doi:10.1515/COGL.2009.022

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00226

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54. doi:10.1037/0033-295X.109.1.35

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, Volume II: The Database* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum.

Marcus, G. F. (2003). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

Matsuo, A., Kita, S., Shinya, Y., Wood, G. C., & Naigles, L. (2012). Japanese two-year-olds use morphosyntax to learn novel verb meanings. *Journal of Child Language, 39*(3), 637–663. doi:10.1017/S0305000911000213

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. doi:10.1037/a0029872

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117. doi:10.1016/S0010-0277(03)00140-9

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393–424. doi:10.1207/s15516709cog2604_1

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. doi: 10.1016/S0010-0285(03)00128-2

Noble, C. F., Rowland, C. F., & Pine, J. M. (2011). Comprehension of argument structure and semantic roles: Evidence from infants and the forced-choice pointing paradigm. *Cognitive Science 35*(5), 963–982. doi:10.1111/j.1551-6709.2011.01175.x

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (Vol. 36, p. 58). Austin, TX: Cognitive Science Society.

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. doi:10.1017/S0305000910000012

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

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua, 92*, 377–410. doi: 10.1016/0024-3841(94)90347-6

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425–469. doi:10.1016/S0364-0213(99)80046-9

Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization. *Cognition, 61*(1–2), 127–159. doi:10.1016/S0010-0277(96)00722-6

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345. doi:10.1111/j.1756-8765.2010.01111.x

Rohde, D. L. (1999). *LENS: The light, efficient network simulator*. School of Computer Science, Carnegie Mellon University. Retrieved from http://www.stanfordlibraries.com/group/mbc/LENSManual/lens.pdf

Rohde, D. L., Gonnerman, L. M., & Plaut, D. C. (2005). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM*, *8*, 627-633.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. doi:10.1016/j.jml.2007.03.002

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition 125*(1), 49–63. doi:10.1016/j.cognition.2012.06.008

Rowland, C. F. & Fletcher, S. L. (2006). The effect of sampling on estimates of lexical specificity and error rates. *Journal of Child Language*, *33*, 859–877. doi:10.1017/S0305000906007537

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. doi:10.1016/j.cognition.2007.06.010

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*(1), 29–56. doi:10.1207/s15516709cog1401_3

St. Augustine (379/2001). *The confessions of Saint Augustine.* C. White (Trans.) London: Frances Lincoln.

Sun, L., & Korhonen, A. (2009). Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2* (pp. 638–647).

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*(1), 127–152.

Tommerdahl, J. (2009). Tommerdahl corpus. *CHILDES*: *Child Language Data Exchange System*. Retrieved December 3, 2012, from http://childes.psy.cmu.edu/browser/index.php?url=Eng-UK/Tommerdahl/

Wells, C. G. (1981*). Learning through interaction: The study of language development*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511620737

Wilson, J. & Henry, A. (1998). Parameter setting within a socially realistic linguistics. *Language in Society, 27*, 1–21. doi:10.1017/S0047404500019709

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. Doi: 10.1016/j.cogpsych.2007.04.002

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. doi:10.1016/S0010-0277(98)00058-4

Wooten, J. (1984). Wooten corpus. *CHILDES*: *Child Language Data Exchange System*. Retrieved December 3, 2012, from http://childes.psy.cmu.edu/browser/index.php?url=Eng-UK-MOR/Wooten/

Appendix A

Table A.1. provides raw frequencies for candidate locative verbs obtained from CHILDES. Because we searched for every possible verb form (e.g., *wind*, *winds*; *pin, pinned, pins*) using the verb stem using wildcards (e.g., wind*; pin*), frequencies in the full corpus include non-verb forms such as *window* and *pink*. Irregular forms (e.g., *wound*) were searched separately. These were excluded from the coded corpus. After coding, this subset contained 1,260 adult and 494 child locative verbs.

Verb	Speaker A	C	Verb	Speaker A	C	Verb	Speaker A	C
adorn	1	0	*glue*	310	104	*smother*	10	0
attach	52	8	*heap*	25	4	*smudge*	39	15
bandage	84	22	*imbue*	0	0	*soak*	76	17
bestrew	0	0	*impregnate*	0	0	*soil*	85	18
bind	46	1	*infect*	15	1	*sow*	7	1
blanket	208	121	*infuse*	0	0	*spatter*	0	0
block	225	118	*inject*	68	16	*speckle*	19	17
blot	0	0	*inlay*	0	0	*spew*	0	0
bombard	0	0	*interlace*	0	0	*spill*	417	153
brush	1160	503	*interlard*	0	0	*spin*	329	137
burden	0	0	*interleave*	0	0	*splash*	418	334
chain	90	23	*intersperse*	0	0	*splatter*	4	3
choke	44	19	*interweave*	0	0	*splotch*	1	0
clog	11	0	*inundate*	2	0	*spot*	229	105
clutter	3	0	*jam*	1202	724	*spray*	66	15
coat	610	287	*ladle*	6	3	*spread*	174	13
coil	4	0	*lard*	6	0	*sprinkle*	65	11
cover	535	217	*lash*	2	2	*squirt*	89	41
cram	11	5	*lasso*	2	0	*stack*	149	28
crowd	42	2	*litter*	23	11	*stain*	34	9
dab	16	8	*load*	389	122	*staple*	1	0
dam	206	67	*nail*	264	109	*stick*	2208	817
dapple	7	1	*occupy*	6	0	*stock*	322	113
daub	0	0	*ornament*	22	4	*streak*	1	0
deck	33	9	*pack*	644	93	*strew*	0	0
deluge	0	0	*pad*	336	75	*stuck*	1553	1272
dirty	1154	579	*paste*	17	5	*stud*	0	0
douse	0	0	*pave*	54	41	*stuff*	553	126
drench	7	0	*pile*	232	56	*suffuse*	0	0
dribble	58	4	*pin*	2440	1182	*tape*	1048	311
drip	124	23	*plaster*	154	102	*tile*	24	4
drizzle	4	1	*plate*	1013	315	*tint*	2	0
dump	311	192	*plug*	120	41	*trim*	50	11
embellish	0	0	*pollute*	0	0	*twirl*	22	9
emblazon	0	0	*pour*	471	158	*twist*	110	25
emit	0	0	*replenish*	0	0	*vein*	1	1
encrust	0	0	*riddle*	1	2	*vomit*	29	12
endow	5	0	*ripple*	2	0	*wad*	12	13
enrich	0	0	*rope*	87	59	*whirl*	5	5
entangle	1	0	*rub*	1217	569	*wind*	2461	788
excrete	0	0	*saturate*	0	0			
expectorate	0	0	*scatter*	15	0			
expel	0	0	*season*	18	0			
exude	0	0	*secrete*	0	0			
face	1620	344	*shake*	280	74			
fasten	98	9	*shroud*	0	0			
festoon	2	0	*slather*	0	0			
fill	357	53	*slop*	35	12			
flood	27	4	*slosh*	8	1			
garnish	0	0	*smear*	5	1			

Appendix B

B.1.1 Layers

The simulations were implemented using Version OSX-1.0a of the LENS connectionist software package (Rohde, 1999; Brouwer, de Kok & Fitz, 2012). Unless otherwise stated, default parameters of the simulator were used. Version of the simulation are available at: http://sites.google.com/site/sentenceproductionmodel/Home/locativemodel.

The sequencing system in the model mapped from Previous Word layer (65 units) to the CCompress layer (10 units) which connected to a Hidden layer (50 units), through the Compress layer (10 units) to the Produced Word layer (65 units). A Context layer (50 units) held a copy of the activation of the hidden layer at the previous time step and connected to the Hidden layer. At the start of each utterance, all Context units were reset to 0.5. The Produced Word layer used the soft-max activation function to create a winner-take-all bias for that layer. The Previous Word layer received one-to-one inputs from all of the Produced Word layer units and from the previous target inputs, and a winner-take-all filter was applied. Thus, during learning from the speech of others, the Previous Word was set to the heard target word. However, when the model was generating its own utterance, the Previous Word layer was set to the word that model had previously produced. This allowed the model to generalise sequencing knowledge learned from heard utterances to its own productions.

The meaning system stored messages in the weights between the Role-Concept bindings (Figure 6, top panel, bottom right), which consisted of the Role layer (4 units) and the Concept layer (60 units). The Hidden layer connected to the Role layer, which connected to the Concept layer, which in turn connected to the Produced Word layer. The model assumed that this "production message" is typically set by the visual scene when learning from others or by the speaker's own message planning during production. The weights between the Role and Concept layers were initially cleared, then for a particular message these Role-Concept bindings between appropriate units were set to a weight of 6 and these weights did not change with learning. Similarly, to allow the model to recognise the role of previously produced words, the model employed a "comprehension message" (Figure 6, top panel, bottom left). This was identical to the production message, except the direction was reversed, mapping from concepts to roles, via weights between the CConcept (60 units) and CRole (4 units) layers. The CRole-CConcept bindings were set simultaneously with the Role-Concept bindings, with a weight of 6.

The Previous Word layer connected to the CConcept layer, which connected to the CRole layer, which in turn connected to the Hidden layer (Figure 6, top panel, bottom left). To ensure the model could avoid producing roles that had already been produced, there was also a CRole Copy layer (4 units; Figure 6, top panel, centre) which averaged a copy of its own activation with the previous activation of the CRole layer. To help the model learn the links between the previous word and its appropriate concept (i.e., the weights between the Previous Word and CConcept layers, Figure 6, top panel, far left), the previous activation of the Concept layer was used as a training signal for the CConcept layer (light grey line, Figure 6). Finally, the meaning system also included an Event Semantics layer (4 units) connected to the Hidden layer. The Role-Concept links in the production message, the CConcept-CRole links in the comprehension message, and Event Semantics activations were all set before a training or test sentence was processed.

Unless specified otherwise, units in all layers used the sigmoidal logistic activation function, with activation values running between 0 and 1. Weights were initially set to values uniformly sampled between –1 and 1. Units were unbiased unless specifically mentioned, in order to make the layers more dependent on their inputs for their behaviour. However, Concept and CConcept units were biased to -3 to ensure that they had a low default activation level.

B.1.2 Training

A version of backpropagation was used to train the model where derivatives were clipped at 1.0 (Doug's momentum; Rohde, 1999). Weights were updated after each message-sentence pair had been trained; the term "epoch" therefore refers to the time taken to train one message-sentence pair. To simulate the gradual reduction in plasticity over development (Johnson & Newport, 1989), the learning rate started at 0.15 and gradually lowered to 0.01 by the end of training. Because in backpropagation models, the error signal becomes weaker as it propagates back through the network, the link between the Previous Word layer, CCompress layer and Hidden layer had a higher, fixed learning rate of 0.1. Training ended after 40,000 sentences had been presented (40,000 epochs) and momentum was 0.9. Since Produced word units had a soft-max activation function, error was measured in terms of divergence between target and produced words as in Equation B1:

(B.1) $\quad \sum t_i \log(t_i/o_i)$

where o_i is the activation for the ith output unit on the current word, and t_i is its target activation due to the softmax function.

Training began by randomising all weights (same model seed was used for all runs). At the start of each utterance, the message was set. After the sentence was generated, the sequence of Produced Word activations was processed by a decoder program that yielded the produced sentence. Sentences were then processed by a syntactic coder program that added the syntactic and message tags. The model's output was compared with the target sentence and the sentence was considered accurate if the all the words were correctly produced.

B.1.3 Input grammars

Inputs to the model consisted of message-sentence pairs generated from a grammar. The grammar included various concepts that were organised into several categories, for example LIVING (MAN, WOMAN, CAT, DOG, BOY, GIRL, FATHER, MOTHER), OBJECT (APPLE, BALL, CHAIR, DRINK, ERASER, FORK, GLUE, HAM), and PLACE (APARTMENT, BOX, CASE, DRAIN, ENVELOPE, FLOOR, GLASS, HOUSE). There were also eight action categories, each containing four action concepts. Five of

these action categories corresponded to locative verb classes (ACTION-A, ACTION-B, ACTION-C, ACTION-D, ACTION-E) one to the intransitive class (ACTION-INT), one to the simple transitive class (ACTION-TRAN), and one to the dative class (ACTION-DAT).

Message-sentence pairs were generated from paired message/sentence templates. The message template specified the set of roles that should be filled and the type of category that should fill that role. For example, the LT locative would have roles for AGENT, THEME and LOCATION, with the AGENT role filled by the LIVING category. To create a specific message, each role was filled by randomly selecting a concept from the appropriate category. For example, when a message stipulated AGENT=LIVING, the AGENT role might be filled with the concept MAN. The sentence template that was paired with the message template specified how the message mapped onto a word order. For example, the LT sentence template specified that AGENT came first, followed by ACTION, then LOCATION, then THEME; each noun was preceded by a determiner; and that the preposition was *with* (Table B.1).

The training grammar included LT and TL locative structures and L and T transitive structures. The locative transitives with verbs from classes B, C and D included only ACTION role information to allow learning of the mapping between lexical semantics of the action and the verb, and there were no other thematic roles or event-semantics. To increase input variability and plausibility, the training grammar also licensed non-locative transitive, intransitive, double object dative and prepositional dative constructions. As locative verbs occurred more frequently in transitive than in locative structures in the corpus analysis, class B – D transitive structures were each four times as frequent as either the class A or E locative structures. Non-locative transitives were 12 times as frequent as class A locatives and other structures were equally frequent as these locatives.

Table B.1. Constructions in the training grammars

LT locative (e.g., class A: LT-only)

Event semantics:	AG LOC TH
Message template:	AGENT = LIVING ACTION = ACTION-A THEME = OBJECT LOCATION = PLACE
Sentence template:	the AGENT ACTION the LOCATION with the THEME
Example sentence:	*the woman fill the glass with the glue*

TL locative (e.g., class E: TL-only)

Event semantics:	AG LOC TH
Message template:	AGENT = LIVING ACTION = ACTION-E THEME = OBJECT LOCATION = PLACE
Sentence template:	the AGENT ACTION the THEME into the LOCATION
Example sentence:	*the woman pour the glue into the glass*

L transitive (e.g., class B: alternating)

Event semantics:	-
Message template:	AGENT = LIVING(omitted in model's message) ACTION = ACTION-B LOCATION = PLACE (omitted in model's message)
Sentence template:	the AGENT ACTION the LOCATION
Sentence example:	*the woman spray the glass*

T transitive (e.g., class D: alternating)

Event semantics:	-
Message template:	AGENT = LIVING(omitted in model's message) ACTION = ACTION-D THEME = OBJECT (omitted in model's message)
Sentence template:	the AGENT ACTION the THEME
Sentence example:	*the woman squirt the glue*

Non-locative transitive

Event semantics:	AG TH
Message template:	AGENT = LIVING ACTION = ACTION-TRAN THEME = OBJECT
Sentence template:	the AGENT ACTION the THEME
Sentence example:	*the woman push the ball*

Intransitive

Event semantics:	TH
Message template:	THEME = LIVING ACTION = ACTION-INT
Sentence template:	the THEME ACTION
Sentence example:	*the woman sleep*

Prepositional dative

Event semantics:	AG LO TH
Message template:	AGENT = LIVING ACTION = ACTION-DAT THEME = OBJECT RECIPIENT = LIVING
Sentence template:	the AGENT ACTION the THEME to the RECIPIENT
Sentence example:	*the woman threw the stick to the dog*

Double object dative

Event semantics:	AG LO TH
Message template:	AGENT = LIVING ACTION = ACTION-DAT THEME = OBJECT RECIPIENT = LIVING
Sentence template:	the AGENT ACTION the RECIPIENT the THEME
Sentence example:	*the woman threw the dog the stick*

Sentences contained no tense or aspect information. Half of the THEME nouns were treated as mass nouns and the article that normally would precede

these nouns (e.g., "the water") was omitted a random 25% of the time. The default preposition for TL sentences was *onto*. However, half of the LOCATION nouns were treated as containers where the preposition *into* was used in TL structures. Critically, although the input grammar generator contained information about these syntactic categories, structures, and verb classes, the model was not given this information. Rather, it developed these representations from the co-occurrence of words in the input distribution.

Training grammars generated 40,000 message-sentence pairs. Since children do not always hear verbs with a matching situation, a randomly-selected 50% of the training inputs appeared without a message. Test sets generated 1,000 randomly-generated message-sentence pairs. The model was trained for at total of 40,000 epochs and tested every 4,000 epochs.