

A distributional learning account of the acquisition of the locative alternation: Corpus analysis and modeling

Katherine E. Twomey (k.twomey@liverpool.ac.uk)

Franklin Chang (franklin.chang@liverpool.ac.uk)

Ben Ambridge (ben.ambridge@liverpool.ac.uk)

University of Liverpool, Department of Experimental Psychology, Eleanor Rathbone Building,
Bedford St South, Liverpool, L69 7WW, UK

Abstract

Early in acquisition children overgeneralize verbs to ungrammatical structures. The retreat from overgeneralization is linked to the acquisition of verb classes, the semantics of which constrain the structures in which a verb can appear (e.g., Pinker 1989; Ambridge, Pine & Rowland, 2012). How children learn these classes remains unclear. Some argue that distributional regularities in linguistic input provide sufficient evidence for verb classes to emerge and become linked to particular structures. A corpus analysis of the English locative construction (e.g., *the woman sprayed water onto the wall/the wall with water*) demonstrated that children have similar verb classes to adults. A correspondence analysis revealed that distributional regularities in the input could support these verb classes. Finally, a connectionist simulation was able to model early overgeneralization and retreat through distributional learning of verb classes. These results support a distributional learning account of verb semantics.

Keywords: Child language acquisition; locative construction; distributional learning; naturalistic speech; Dual-Path model.

Children's acquisition of grammatical constructions remains one of language acquisition's most puzzling phenomena. Even young children are readily able to generalize known constructions to new verbs, and indeed, early in acquisition children overgeneralize these constructions to contexts in which they are ungrammatical (e.g., **she filled the juice into the glass*). However, over time children learn to constrain their choice of structure, despite the lack of evidence for ungrammaticality in the adult speech they hear (Ambridge, Pine, Rowland, Jones & Clark, 2009).

This remarkable ability has been linked to the semantics of the verbs that appear in these constructions. According to this semantic verb class hypothesis (Pinker, 1989), verbs fall into discriminable classes based on the semantics of the event they describe. These verb classes then constrain the structures in which verbs can appear. For example, *pour* and *drip* both describe the movement of an object/substance into/onto a location via gravity, and both appear in the theme-locative (TL) structure only (e.g., *the woman poured water into the bucket* vs. **the woman poured the bucket with water*). In contrast, *cover* and *coat* both describe a location being completely covered with a layer of substance or object(s) (e.g., *the woman covered the table with the blanket* vs. **the woman covered the blanket onto the table*). Finally, *spray* and *squirt* both describe ballistic motion of a

liquid along a trajectory, are in a class that does not emphasize theme or location, and therefore alternate between both LT and TL structures (e.g., *the woman sprayed the wall with water* vs. *the woman sprayed water onto the wall*). Although Pinker's verb classes motivated early work on such verb-structure mappings, recent work has questioned whether it is these particular classes which guide structure choice (Ambridge, Pine & Rowland, 2012). Nonetheless, it is widely agreed speakers categorize verbs, and that these classes help to guide structural choices and explain overgeneralization.

One account of the acquisition of semantics is based on using situational information in the world (St. Augustine, 397/2001). For verbs, this might involve learning that verbs like *coat* refer to events where an object is completely covered with some substance. Since the object is more saliently changed by the action than the substance, the situational meaning would classify this verb as belonging to a verb class that prefers the LT structure, which emphasizes the LOCATION (e.g., *the man coated the [table]_{LOCATION} with paint*). However, this situational account of verb meaning is limited: it may not be the case that every child hears every verb that they know in an unambiguous situational context that classifies it precisely (e.g., Harris, Jones & Grant, 1983).

An alternative, distributional account states that verb semantics are learned from distributional regularities in the linguistic input (Fisher, Gleitman, & Gleitman, 1991). For example, if *spray* is always followed by a noun phrase which specifies a liquid (e.g., *spray the water*), then children might categorize *spray* with other verbs that take similar arguments (e.g., *splash the water*). This verb class might then be associated with certain locative structures (e.g., *he sprayed/splashed the water onto the wall*). The advantage of this account over the situational account is that it allows verb classes to be acquired from heard input without any situational context.

To examine how these accounts relate to behavioral data, we performed a corpus analysis of the British English locative. Then, to explore whether distributional information is sufficient to support the emergence of verb classes, we performed a correspondence analysis on the corpus data. Finally we investigated whether distributional regularities can influence structural choice in a connectionist model of sentence production which learns both syntactic

structures and their verb class associations from the input alone.

A corpus analysis of the English locative

We queried all UK corpora in CHILDES (MacWhinney, 2000) for the 142 locative verbs listed in Pinker (1989). A main sample of 38,231 utterances was retrieved. A subsample of 2,685 morphologically-tagged utterances was coded by hand for construction type (see Table 1). Importantly, we divided transitives into those where the post-verbal noun was a THEME (*spray the water*, T transitive) or a LOCATION (*spray the wall*, L transitive).

Table 1: Constructions and handcoded frequency

Structure	Example	N
LT locative	brushing me with it?	100
TL locative	spread it on your biscuit	232
L transitive	for him filling the glasses	271
T transitive	you dump the lady's toys	294
Ambiguous transitive	brushed it	608
L intransitive	tape on	24
T intransitive	it attached on like that	20
Ambiguous intransitive	I brushing	205

The subsample was generated by selecting all utterances containing candidate locative verbs tagged as a verb or a participle. Utterances were separated by speaker (adult or child) and post-verbal preposition (*with*, indicating a candidate LT locative; *into*, *onto* or *over*, indicating a candidate TL locative; or no preposition, indicating some other construction). For the 13 categories containing more than 50 utterances, a random sample of 50 utterances was coded. Since there were very few tagged utterances for conversations involving children over 40 months, we extracted an extra 20% of child utterances and 5% of adult utterances from the untagged data for this age group. We excluded 221 non-verb utterances from the analyses. As can be seen in Table 1, only 6% of the coded locative verbs occurred in the LT locative structure, and 13% occurred in the TL locative structure.

The situational account predicts that verb classes in children and adults can differ, especially if children do not always understand events in adult-like ways. On the other hand, the distributional account predicts that children learn their verb classes from adult linguistic utterances, so there should be a close match between the two. To examine these predictions, we calculated the proportion of LT out of total locative constructions in the hand-coded data. This controls for raw frequency of each form and variation due to other structures, providing a measure of the bias for these two structures. Figure 1 depicts proportion LT for both children and adults for a range of verbs that occur more than 10 times in the hand-coded adult data. There seems to be a class of non-alternating LT verbs like *fill*, a class of non-alternating TL verbs like *pour*, and several alternating verbs

like *rub* and *splash*. Importantly, children have similar verb classes to adults, even though these data come from a range of different situations, which children may or may not understand in adult-like ways.

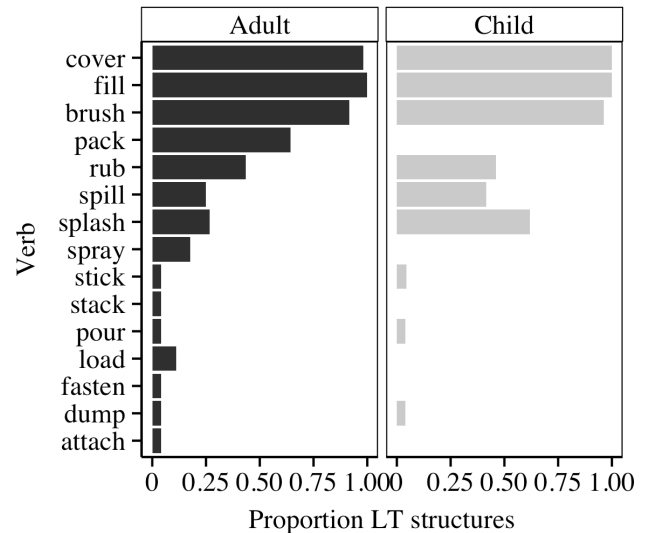


Figure 1: Proportion LT structures, adult and child speakers

To see whether these structural preferences change over development, we calculated proportion of adult and child transitive and locative utterances with location as object (L or LT) for each age of child included in the sample (Figure 2). A linear model was fit to the data with Structure (locative = 1, transitive = 0), Child Age (months) and Speaker Group (adult vs. child) crossed. There was no main effect of age, suggesting that L transitive proportion did not vary over age. LT production was lower than L transitive production (beta = -0.9, $t(116) = 3.1$, $p < .003$). An interaction between Age and Structure (beta = 0.02, $t(116) = 3.1$, $p < .003$) provided evidence for an increase in the production of LT structures over development. The lack of any interaction with Speaker Group suggests that adults adapted their structural choices to fit children's preferences. The TL bias in young children is also seen in experimental studies, where children reproduce TL structures at a higher rate than LT structures (Gropen et al., 1991a, 1991b), and in diary studies, where TL overgeneralizations are more frequent (Bowerman, 1982).

In our data set, transitives are more frequent than locatives – we found 565 transitive utterances, but only 332 locative utterances. Of the 50 verbs that appeared in a transitive, a locative or both structures, 35 appear in the locative while 44 appear in the transitive. Using transitive contexts to learn about locative verbs provides a better coverage of these verbs, suggesting that children may be able to learn about locative verbs from their appearance in L and T transitive structures.

To examine this hypothesis, we used adult ratings of locative verbs' LT and TL preference from Ambridge et al. (2012) and Bidgood, Ambridge, Pine & Rowland (under

review) to create a graded LT rating measure for each verb. This was correlated with the proportion LT and L structures in the adult input to children from the corpora. There was a significant correlation between the LT rating and LT corpus distribution ($r(30) = .53, p = .002$), indicating that the LT/TL biases of verbs in locative structures in child-directed speech match adult ratings of those verbs' LT/TL preference. There was also a significant correlation between LT rating and L corpus distribution ($r(38) = .32, p = .04$), which shows that the L/T biases of verbs in transitive structures in child-directed speech also reflect LT/TL preference ratings.

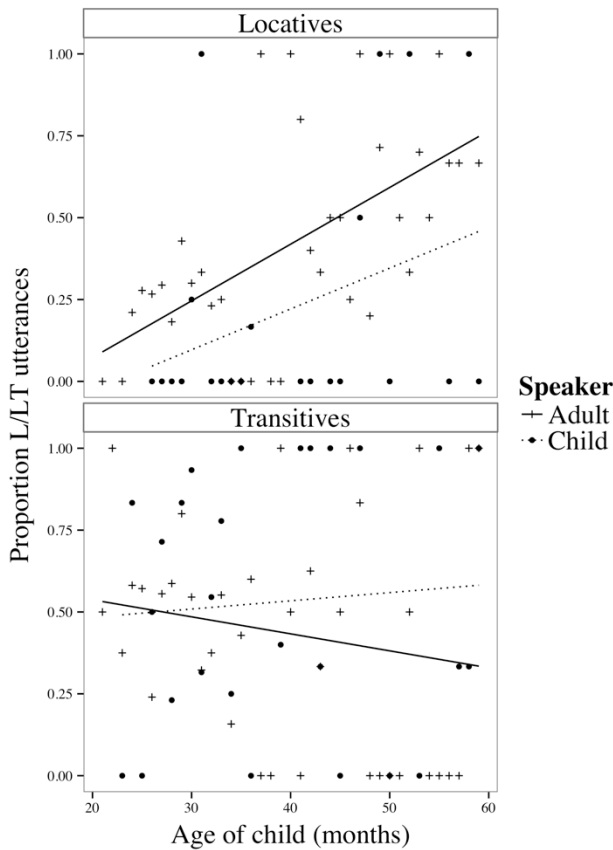


Figure 2: Proportion L/LT locatives by child age

Correspondence analysis

The corpus analysis shows that children learn biased verb classes in their first five years. We used correspondence analysis (CA; Greenacre, 2007) to investigate whether distributional regularities can support the development of these classes. CA is a technique for clustering categorical data in a low-dimensional space based on the similarity in how elements co-occur.

We performed a CA on adult utterances including utterances that were not hand-coded. Similarity was calculated based on the co-occurrence of each verb with the two post-verbal words, however if a preposition (*on, to, with, in, into, onto, over*) occurred post-verbally, the

preposition and all following words were excluded. Thus, verbs that occurred in the same context (e.g., *the woman poured water* and *a boy dripped water*) were classified as more similar than verbs that occurred in different contexts. In order to have a range of words to support categorization, we only included verbs which occurred more than 40 times. The CA generated six factor scores for each verb.

To determine which factors in the CA best predicted the LT/TL rating measure used earlier, we submitted factor scores for each verb to a regression with LT rating as the dependent variable, and Factor (1, 2, 3, 4, 5, 6) as independent variables. LT rating was significantly predicted by Factor 4 ($\beta = 0.03, t(27) = 2.06, p = .05$) and Factor 6 ($\beta = 0.06, t(27) = 3.72, p < .001$), explaining 39% of the variance ($F(6,27) = 2.85, p = .03$).

Figure 3 illustrates the relationship between Factors 4 and 6 and LT rating. Verbs rated as more grammatical in LT structures (e.g., *the woman poured water into the bucket*) cluster in the top right hand corner of the figure, and verbs rated as more grammatical in TL structures (e.g., *the woman filled the bucket with water*) cluster in the bottom left hand corner. Furthermore, there is evidence of clusters reflecting Pinker's (1989) narrow semantic verb classes; for example, *fill* and *cover* are both members of Pinker's *cover*-type class and they are clustered together. In line with existing studies (e.g., Ambridge, Pine & Rowland, 2012), however, not all classes generated by our data conform to these narrow verb classes. For example, *fill* and *pack* are similar on Factor 4, despite being members of different classes according to Pinker's classification.

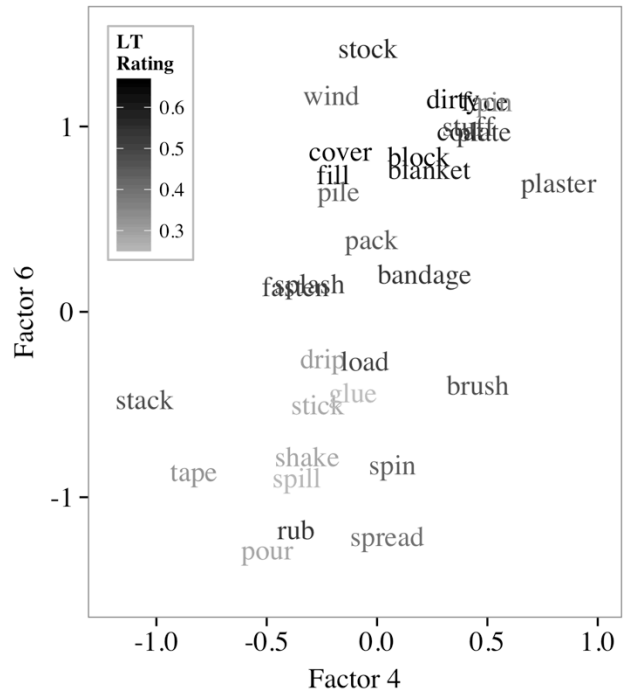


Figure 3. CA results for Factors 4 and 6.

Overall, the distributional regularities with which verbs and post-verbal words co-occurred generated two factors in a CA which predicted adult ratings of verbs’ grammaticality in LT structures. Substantial evidence exists that children can track such statistical co-occurrences in their environment (Fisher, Gertner, Scott & Yuan, 2010). Thus, if children record the distributional regularities of locative verbs, they may also create verb classes, helping them to constrain their choice of structure and retreat from overgeneralization. The CA provides an explicit model of how a distributional learning mechanism can create these classes.

A connectionist model of locative acquisition

The corpus analysis suggested that children may learn about the semantics of locative verbs from transitive structures. Further, the correspondence analysis showed that distributional regularities in the input could support the emergence of verb classes that constrain structural choice. However, a full account requires a mechanism that can learn not only verb classes but also syntactic structures, and link them appropriately. Critically, if this mechanism is similar to the language acquisition system in children, then it should create a TL bias early in development.

To explore the nature of such a mechanism, we adapted the Dual-Path model, a connectionist model of sentence production (Chang, 2009, Chang, Dell & Bock, 2006). The model learns linguistic representations from message-sentence pairs and can acquire different languages (e.g., English, Japanese). We trained the model with two simple grammars that included both transitive and locative structures. Grammars contained five classes of verbs, one of which was L/LT -only and one of which was T/TL -only. Our principal manipulation was the frequency with which the remaining “alternating” verb classes were associated with the various structures.

Model

The Dual-Path model’s architecture includes separate sequencing and meaning systems. The sequencing system is a simple recurrent network that allows the model to learn syntactic representations (Elman, 1990). The model’s internal representations are acquired through error-based learning, in which the model predicts the next word in a sentence, then uses the difference between the prediction and target (error) to modify its internal representations. Since categories are useful for prediction, the sequencing system acquires syntactic categories that support syntactic structures. Verb classes can be seen as a refinement of verb categories to incorporate their biases. The meaning system encodes the message as a set of weights between role and concept units (e.g., AGENT=WOMAN). The simple recurrent network selects appropriate concepts by activating their role at particular positions in sentences. Detailed motivation for the model’s architecture is provided in Chang (2002; 2009).

Messages include roles for the various concepts in the sentence as well as a role that selects the verb. In the

current architecture, structural alternations were modeled by associating different structures with the same message. For example, the TL locative sentence *the woman sprays water onto the wall* and LT locative sentence *the woman sprays the wall with water* have approximately the same meaning and hence share the message ACTION=SPRAY AGENT=WOMAN THEME=WATER LOCATION=WALL. When given this message, the model must decide whether to activate to LOCATION role after the verb and produce the LT structure, or the THEME role and produce the TL structure. Critically, because they are located in the meaning system, the SPRAY semantics can only be used for selecting the word *spray* and do not directly influence structural choices in the sequencing system. Thus, the model must acquire a syntactic representation of the verb within the sequencing system using distributional learning.

Testing different input distributions

Verb classes in the simple recurrent network were shaped by the distributional properties of the input. To see if these classes reflected the behavior of human children’s verb classes, we compared two different input distributions: Full and Transitive (Table 2).

Table 2: L/LT ratios in input grammars (LT:TL)

Structure	Full Distribution	Transitive Distribution
Locative A	100:0	100:0
Locative B	90:10	-
Locative C	50:50	-
Locative D	10:90	-
Locative E	0:100	0:100
Transitive A	100:0	100:0
Transitive B	90:10	100:0
Transitive C	50:50	50:50
Transitive D	10:90	0:100
Transitive E	0:100	0:100

Our input distributions contained five classes of verbs (A, B, C, D, E), each paired with LT and/or TL locative structures (e.g., *the woman sprayed the wall with water / water onto the wall*) and L and/or T transitive structures (e.g., *the woman sprayed the wall/water*). Verbs in class A occurred only in L/LT structures, while verbs in class E occurred only in T/TL structures. In the “Full Distribution” simulation, verbs occurred more frequently in L/LT structures in Class B, equally often in either structure type in Class C, or more frequently in T/TL structures in Class D. The “Transitive Distribution” simulation tested the idea suggested by the corpus and correspondence analysis that transitive uses of verbs might be the basis for verb classes. Classes B, C, and D were therefore only biased in transitive structures (and more strongly than in the Full Distribution) and hence tested whether the transitive distribution could influence locative generalization. The raw frequency of TL structures was also higher than that of LT structures in the Transitive Distribution to simulate the data in Table 1.

Ten randomly generated training sets were used to create ten model subjects. Each model was trained for 40,000 epochs on 2,000 message-sentence pairs. Every 1,000 epochs the model was tested on a set of 1,000 grammatical and ungrammatical locative sentences (100 per verb class).

Results. Figure 4 depicts proportion location-based sentences for each of the five verb classes, for each of the simulations.

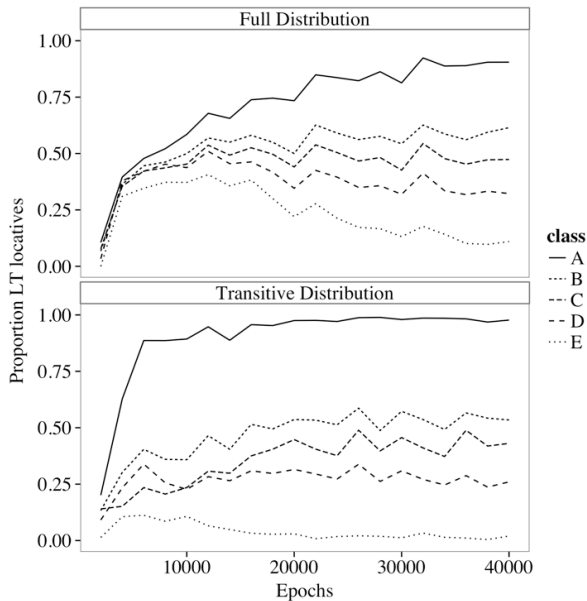


Figure 4: Proportion location-based sentences

Results from the Full Distribution (Figure 4 top panel) suggest that distributional regularities in the input do indeed support the emergence of verb classes. First, the Full Distribution simulation learned nonalternating classes A and E. To establish whether the model distinguished classes B, C and D according to the frequency of L/LT and T/TL structures for these classes in the input, a mixed effects model was fit to Proportion LT (empirical logit transformed) with Verb Class and Epoch crossed. Verb Class was coded numerically (B = 1, C = 0, D = -1). Model subject was included as a random variable and there were by-subject slopes for Verb Class crossed with Epoch (Barr, Levy, Scheepers & Tily, 2013). There were no main effects of Verb class or Epoch on LT production, however an interaction ($\beta = 0.016$, $\chi^2(1) = 14.49$, $p < .001$), confirmed that verb classes differentiated as epoch increased. The negative logit intercept coefficient captures the model's initial TL bias ($\beta = -0.16$).

The Transitive Distribution simulation (Figure 4 bottom panel) also distinguished classes B, C and D in the locative, despite receiving no locative input for these classes. It did so based on the frequency with which these verb classes occurred in L and T structures. Class D, for example, was a class that preferred THEME-type objects rather than

LOCATION-type objects in the post-verbal position. To test this, we fitted a mixed effects model to the Transitive Distribution data. There were no main effects of Verb Class or Epoch on LT production, but again there was an interaction ($\beta = 0.01$, $\chi^2(1) = 6.90$, $p = .009$), demonstrating that verb classes differentiate over development. Finally, this model also exhibited an early TL bias, as denoted by the negative logit intercept coefficient ($\beta = -0.54$). The preferences in the transitive naturally generalize to locative structures because these structures have similar argument preferences (i.e., TL locatives prefer theme-type objects in the post-verbal position). Thus, the model is able to acquire locative verb classes from transitive distributions.

Our corpus analysis revealed that children have an early preference for the TL locative. This was captured by the model. Like the children, LT production increases over development, but remains outstripped by TL production (overall proportion LT for Transitive Distribution at epoch 10,000 = 0.36, epoch 40,000 = 0.44; Fig. 4). This simulation captures the early TL bias because the sequencing system does not have access to verb meaning and is biased to learn syntactic categories that fit the frequent TL structure. This structure becomes the default means to express three arguments (i.e., AGENT, THEME, LOCATION). Slowly the model develops verb classes that help predict the nouns that follow verbs. These classes later help support the differential use of LT and TL structures.

General Discussion

The current study presents naturalistic data which show that children are biased to use TL structures early in acquisition. We hypothesized that the more frequent TL overgeneralization errors in diary studies are due to this general TL bias. The increase in use of LT between two and five years can therefore be interpreted as a metric of an underlying retreat from overgeneralization. If, over development, children acquire verb classes that predict which structure to use, then they will become more likely to produce these structures in verb-specific ways, and consequently less likely to make errors. To explain how children acquire verb classes when the target classes and structures are not explicitly provided, we suggest that they perform a distributional analysis on the words they hear, in a similar manner to our CA. Here, we have shown that an unsupervised clustering algorithm can create component factors that predict structural choices.

Bayesian and other statistical mechanisms have also been applied to verb-structure learning. These systems often assume that the initial state of the learner includes syntactic structures, semantic features (e.g., cause, change-of-state), or even abstract verb classes (Alishahi & Stevenson, 2008; Niyogi, 2002; Perfors, Tenenbaum & Wonnacott, 2010). A clear example of this approach can be seen in Niyogi (2002), who developed a Bayesian model of the locative alternation. The model uses manner and path features from scene and syntactic frame regularities to select among verb

class hypotheses. A key feature of this model is that it can assign verb classes quickly from a few exposures; it therefore does not explain the protracted development of the LT structure in children. Furthermore, the model cannot learn locative classes from transitive input, because locative verb classes are not an optimal fit for transitive structures (superordinate motion classes are a better fit). The limitations of these models highlight the fact that language development is not just the fast, optimal weighting of syntax and semantic cues.

In contrast with these models, the Dual-Path model does not start with syntactic structures or verb class hypotheses. It learns its structures with a slow learning algorithm designed to mimic the slow biological changes that support learning in the brain (cell growth). Due to its inability to predict when to use LT and TL structures, the frequent TL initially dominates. As in the CA, the model develops verb classes from the distribution of post-verbal words in frequent transitives. Later, these verb classes become the basis for distinguishing TL and LT structures and thus condition their use.

Taken together, the current studies are the first to characterize how children use a wide range of verbs in locative structures between two to five years of age. We showed that locative production is asymmetrical during this period, with the TL structure dominating. This dominance can cause children to place a newly-learned verb into the TL structure regardless of its own bias, producing an overgeneralization error. We also showed that distributional regularities in child-directed speech are useful for creating classes that predict adult ratings of locative structure preference for this diverse set of verbs. Finally, we showed how distributional learning in the Dual-Path model can be used to learn both the syntactic categories that support structures, as well as the verb classes that bias structural choice. By trying to simultaneously solve both of these difficult learning problems, the model can explain both the early TL bias and slow development of verb classes that support the retreat from that bias.

Acknowledgments

This research was supported by Leverhulme Trust Research Project Grant RPG-158.

References

- Alishahi, A., & Stevenson, S. (2008). A Computational model of early argument structure acquisition. *Cognitive Science*, 32(5), 789–834.
- Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, 123(2), 260–279.
- 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.
- 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.
- Bidgood, Ambridge, Pine & Rowland (under review). The retreat from locative overgeneralisation errors: A novel verb grammaticality judgment study.
- 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 319-346). New York: Cambridge University Press.
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26(5), 609–651.
- 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., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Fisher, C., Gertner, Y., Scott, R., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(2), 143-149.
- Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, 23(3), 331–392.
- Greenacre, M. (2007). *Correspondence analysis in practice*. Boca Raton, FL: Chapman & Hall/CRC.
- Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991a). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language*, 18(1), 115–151.
- Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991b). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, 41(1), 153–195.
- Harris, M., Jones, D., & Grant, J. (1983). The nonverbal context of mothers’ speech to infants. *First Language*, 4(10), 21–30.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum.
- 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)
- Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37, 607–642.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: The MIT Press.
- St. Augustine (379/2001). *The confessions of Saint Augustine*. C. White (Trans.) London: Frances Lincoln.