Linguistic input overrides conceptual biases: When goals don't matter

Nicholas A. Lester (nlester@umail.ucsb.edu)
Department of Linguistics, South Hall 3432
Santa Barbara, CA 93106 USA

Abstract

Previous research has suggested the presence of a cognitive *goal bias*, which favors the linguistic and non-linguistic (mnemonic) encoding of goals of motion over sources. The present corpus-based study tests the limits of the goal bias by comparing the path-encoding tendencies of English *come* and *go* in young children's and adults' naturalistic speech. Both verbs can occur with source and goal adjuncts; however, they differ in their presuppositional structure, such that *come* presupposes a goal while *go* presupposes a source. This difference in presupposition might lead adult speakers to inhibit goal-encoding for *come* via the Gricean maxim of quantity. As input, this might lead young children to acquire different path structures for *go* and *come*, even before they have mastered conversational-pragmatic abilities. Descriptive statistics replicate earlier findings of a general goal bias for both verbs. However, the results of more detailed regression analyses suggest that *go* exhibits a stronger goal bias than *come* for children and adults. Moreover, children from ages 2-3 persist in inhibiting goal mentioning for *come* at rates similar to adult usage. This effect holds even while goal expressions for *go* are becoming more complex. These findings suggest that statistical patterns in the input can override non-linguistic biases, even during early lexical acquisition.

Keywords: Goal bias; child language; usage-based grammar; motion deixis.

Introduction

Understanding how children acquire language requires in part an understanding of how linguistic forms come to be associated with non-linguistic conceptual content during word learning. Some researchers have argued that children accomplish this task by exploiting *conceptual homologies*, or isometries in the representations and behaviors of non-linguistic and linguistic systems. These homologies aid children by splitting the conceptual landscape into ready-made sites for the mapping of linguistic forms. The relationships established by these homologies resemble the *linking rules* proposed by Jackendoff (1990), who would argue further that they are available from birth. One homology which has figured prominently in recent debates is known as the *goal bias*. The goal bias constitutes a domain-general attentional preference for encoding the end-points of motion as opposed to the sources. In language, one consequence of the goal bias is that goal landmarks are mapped by default into the adjunct prepositional phrase (PP) of motion predicates. By contrast, source landmarks in these predicates are only usually mentioned in conjunction with a goal landmark or not at all. The goal bias could thus guide children's acquisition of motion verbs by providing them a point of constancy for the interpretation of path-relevant adjunct PPs and adverbials.

In order for the goal bias to prefigure mapping relations during word learning, it must operate from a very early age. Support for early appearance of the goal bias comes from the fact that pre-linguistic infants as young as 12 months old have shown a preference for attending to changes in goals over changes in sources (all else being equal) in preferential looking paradigms (Lakusta, Wagner, O'Hearn, & Landau, 2007). This effect persists in memory tasks for older children (3-7 years old) and adults (Lakusta & Landau, 2005; Lakusta & Landau, 2012; Papafragou, 2010), and so appears to be a stable component of the human cognitive apparatus.

The simplest demonstration of a linguistic goal bias in English comes from the fact that goals tend to be mentioned more frequently than other path components in naturalistic speech. In the only major corpus study to examine motion verbs specifically, Stefanowitsch and Rohde (2004) found higher overall rates of goal-mentioning than source mentioning across a number of motion verbs in a corpus of journalistic writing. Similar findings have been reported from experimental elicitation tasks, in which both children and adults were more likely to mention goals than sources when prompted to describe short video clips of motion events with equally salient goal and source landmarks (Lakusta & Landau, 2005; 2012; Papafragou, 2010).

More recent evidence suggests that the goal bias is more robust for language than for memory, drawing the notion of a strict isometry into question. Lakusta and Landau (2012; Exp. 3) showed participants motion events involving the incidental motion of inanimate *themes* (moving entities). Participants' memory for contrasts was equally reliable for goals and sources. However, when asked to describe the events, they remained more likely to refer to goal landmarks than to source landmarks. This finding held for both adults and children.

Some evidence, however, suggests that the goal bias may be overridden in certain contexts for language and memory. For example, when Stefanowitsch and Rohde (2004) looked at goal-mentioning rates for individual verbs (instead of the aggregate values), they found a great deal of variation, with some verbs showing no goal bias at all. But what could drive such variation? Stefanowitsch and Rohde attribute it to a combination of lexical semantics and pragmatic knowledge. Lakusta et al. (2007) suggest that situational factors may play a role. They found that by increasing the visual salience of source landmarks within a motional scene, they could eliminate the non-linguistic goal bias effect in 12-month-olds. Therefore, children appear able to shift attention between goals and sources in response to

contextual pressures. These shifts could account for the variation observed by Stefanowitsch and Rohde if different verbs become associated with situations involving low-salience goals or high-salience sources. Lakusta and Landau (2012) speculate about the potential sources of these attentional shifts beyond perceptual asymmetries. They suggest that presupposition may be a factor, such that presupposed path elements will not be mentioned (a la the Gricean maxim of quantity). This possibility is also suggested by Stefanowitsch and Rohde, who attribute their observed source bias for *fall* to presuppositions about the nature of gravity.

Lester (2014) tested the prediction that lexical (as opposed to contextual) presuppositions influence path mentioning by comparing adults' use of source and goal adjuncts with the English deictic motion verbs *come* and *go*. *Come* has been analyzed as presupposing the goal of motion, inasmuch as the goal must be construable as an extension of the speech act's deictic center or 'here' space (e.g., MOVE (x, TO (HERE)). *Go*, on the other hand, presupposes the source of motion (e.g., MOVE (x, FROM (HERE); Talmy, 2000). In line with predictions, Lester found that adults did exhibit a significantly more pronounced goal bias for *go* than for *come*.

Another point of controversy concerns whether the goal bias operates the same in children as adults. Lakusta and Landau (2012) predict that children should mirror adults in their goal-mentioning. In their words, "'proto-goals' and 'proto-sources' for children may be quite similar to those for adults, and thus there will be few developmental differences in how sources and goals are mapped into syntax" (p. 24). However, their own findings show that in some cases, the goal bias persists in child language (and to some extent, memory) where it disappears for adults (Exp. 2). In fact, the children showed a goal bias in every linguistic task reported in that study. Therefore, it may be that children are more susceptible to the goal bias than adults, in which case they should start out by mapping goals exclusively (regardless of motion verb). This hypothesis also meshes with theories of language acquisition which specify innate linking rules responsible for ensuring successful mapping of phonological form to conceptual unit (e.g., Jackendoff, 1990). Yet another possibility is that children's path mentioning is a direct reflection of the patterns contained in the linguistic input. This hypothesis makes the same ultimate predictions as Landau and Lakusta's 'same-system/same-behavior' proposal, but has the advantage of explaining the source of possible variations in goal bias (like those observed by Stefanowitsch and Rohde, 2004).

In the next section, I formalize several predictions regarding the effects of lexical presupposition on children's developing path encoding skills and the extent to which these effects can be attributed to general biases, patterns in adult speech, or a combination of both.

Hypotheses

If children indeed leverage the goal bias during the learning of motion verbs, then we should expect them to produce more goal mappings for all verbs in the early stages of acquisition. We should expect this because the child would not have any reason to guess against the goal bias when selecting between a source- or goal-interpretation of a given verb-landmark combination, and so should not yet be sensitive to verb-specific variations in such pairings. Regier (1996) makes the same prediction, but from a different theoretical perspective. He suggests that goals are perceptually and mnemonically more accessible in situated events of language use. As children puzzle out the meaning of a motion verb that has just been uttered, they might more readily associate that verb with the goal configuration due to a kind of recency effect in working memory (the goal having been the last relevant perceptual input). In either case, the prediction can be formulated as in H1:

H1: Children should exhibit an initial goal bias for *come* and *go*, but the former should decrease over time while the latter should increase or stay constant

As a counter to H1, consider the input-driven position that language use shapes conceptual representation. According to this position, statistical patterns in the linguistic input should dictate children's path-encoding, regardless of the goal bias. This prediction is presented as H2:

H2: Children at even the earliest stages of syntactic development should exhibit a goal bias for *go* but not for *come*

H2 would receive the strongest support from a correlation between goal bias and verb identity with no corresponding interaction of verb and age. This would indicate that verbs can differ in goal bias but that children's path mentioning is identical to adults' path mentioning. In the more realistic scenario, verb identity will interact with age, though the sign of the difference in goal bias between *come* and *go* in the child data should match that of the adult pattern. This result might suggest the presence of more general constraints on the length and complexity of children's spontaneous utterances, though the utterances do approximate the adult usage.

Data

The child data for this study were culled from the Manchester subcomponent of the CHILDES database (Theakston et al., 2001). The Manchester corpus contains approximately 2.1m words of transcribed recordings of naturalistic speech, including 580,472 words of child speech from 12 children (n(female) = 6; figures adopted from Li & Fang, 2011). Children's initial ages in the sample ranged from 1;8.22 to 2;0.25. MLU estimates ranged from 1.06–2.27. The data capture spontaneous interactions arising during free play activities with the child's own toys or with

a special set of investigator-supplied toys. One-hour recordings were taken twice in every three-week period over the course of one year. This sample was selected for several reasons. First, it constitutes a relatively dense sample, which is crucial for establishing realistic developmental trajectories. Second, it captures the earliest stages of syntactic development and so is well suited to test for the heightened sensitivity to the goal bias predicted by H1. Third, the activities from which the speech was sampled naturally lend themselves to increased motion verb use (playing with trains, cars, etc.). Fourth, we know from prior research that children at these ages have not mastered the complexities of deictic reference and presupposition (Clark & Garnica, 1974). Therefore, any differences in path mentioning cannot be attributed to the children formulating but omitting goals in conformity to the maxim of quantity.

I retrieved all instances of all inflectional variants of *come* and *go* from the child corpus. To account for the possibility of non-standard forms in the child speech, (e.g., overgeneralization errors such as *comed* or *goed*), I relied on the Manchester corpus' error annotation, which conveniently includes tags for the target (standard) form after each non-standard productionThis search yielded 1975 tokens of *come* and 5043 tokens of *go*.

I next removed all false hits from the child data. This mainly consisted of removing locative and affiliative uses of *go*. Locative uses consisted of instances of (some identifiable fragment of) the fixed expression *Where has* X *gone*? These instances can be paraphrased by '*Where is X?*' Therefore, while they may impose enough of a motional construal to exhibit effects similar to 'fully' motional tokens, I omit them here for the sake of simplification. Affiliative uses of *go* were more tricky to identify, and so were labeled more conservatively. In an affiliative function, *go* expresses the proper relation between an entity and a location, as in the expression *The salt goes in the cupboard.* The children in this sample often used *go* in this way when orienting objects within the play-space (e.g., *This barn goes here*) or when expressing properties of the objects to the interlocutor (e.g., *Cows go in the barn*). Again, as the motional status of these utterances is suspect, they were omitted from the present analysis. I also removed any uninterpretable utterances (utterances with ### indicating that the transcriber was unable to determine what was said; truncated utterances, as in *He went in the*) and any complex clauses involving *come* or *go* (as in *Come get this* or *Go and shop in town*). This process left me with 1860 tokens of *come* and 4123 tokens of *go.*

For the adult baseline, I took the data from Lester (2014), which consisted of a proportional random sample of all literal translation motion uses of *come* and *go* as retrieved from the roughly 3.2-m-word spoken component of the Open American National Corpus (OANC). The spoken OANC consists of a sample of the SWITCHBOARD corpus (telephone conversations on set topics between strangers; approx. 3 m words) and the entirety of the Charlotte Narrative and Conversation Collection (face-to-face conversations and conversational narratives; approx. 198,000 words). This sample contains 454 instances of *come* and 1417 instances of *go.*

I annotated the data for a number of variables characterizing features of the theme, source, trajectory, and goal. In addition, I annotated each instance with a speaker identification label (child's name for Manchester data; OANC filename for the adult data) in order to account for the non-independence of the data points.

Descriptive stats

Overall relative proportions of mentioning for goal and source landmarks, ignoring their relative distribution within clauses are presented in Figure 1.

As has been observed in previous studies, Figure 1 reveals that a general goal bias (i.e., collapsing across verb types and ages) holds in the form of a greater overall token frequency of goals as opposed to sources. Figure 1 also shows a relatively strong goal bias for *go* in both child and

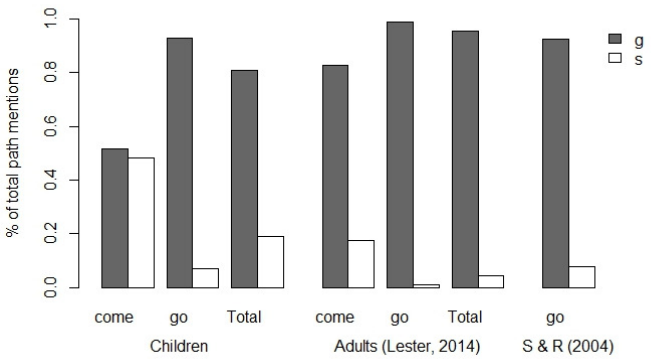

Figure 1: Relative proportions of use of goal (g) and source (s) landmarks

adult speech, with numbers consistent with those reported in Stefanowitsch & Rohde (2004; shown in the final column of Table 1). However, for *come*, the adults exhibit a weaker goal bias than *go,* and the children exhibit no goal bias whatsoever. This latter finding is particularly surprising given the logic underlying H1: children are not influenced more heavily by the goal bias but actually seem to resist it when the adult usage presents the potential for a meaningful contrast (i.e., the fact that *come* is 16 times more likely than *go* to appear with a source). In the following two studies, I attempt to explore these verb-specific effects in more detail. Study 1 focuses on the child data, looking within the longitudinal scale of the Manchester sample to test the strength of the goal bias across early periods of acquisition (H1). Study 2 compares the aggregated child data (ignoring the Manchester-internal time-scale) to the adults to determine whether differences in child path-mentioning can be attributed to adult usage (H2). For both studies, I operationalize goal bias as the difference in length (in orthographic characters) of any source landmark NP from that of any goal landmark NP within the same clause. When a clause did not contain a source or goal landmark, a value

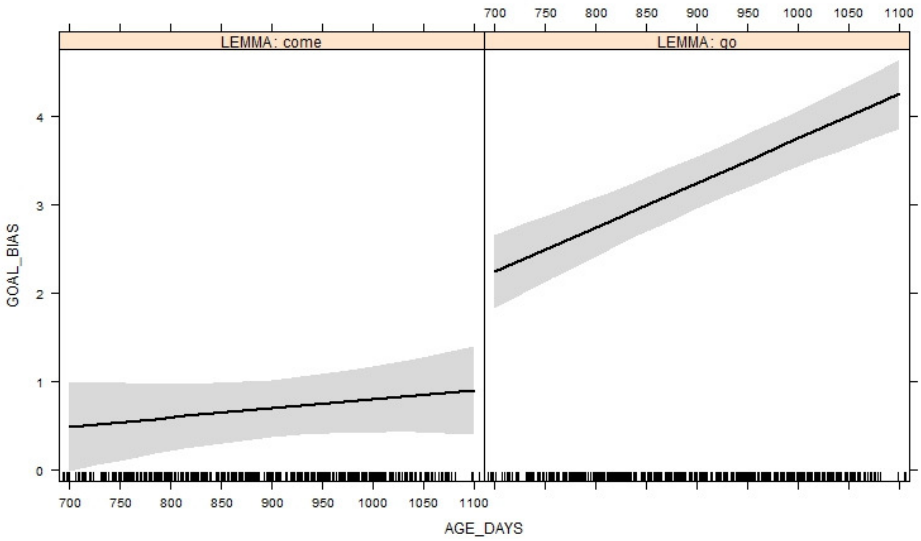

Figure 2: Goal bias as a function of the interaction of speaker age (in days) and verb (*come, go*)

of 0 was assigned. Positive values indicate a goal bias, while negative values indicate a source bias within a given clause. This operationalization improves on that used in previous analyses (e.g., Stefanowitsch and Rohde, 2004) by going beyond simple frequency of mentioning to provide an estimate of the relative complexity of the path expressions.

Study 1: Path-mentioning in child speech

As a first step, I investigate whether children are more susceptible to the goal bias, as suggested by the data in Lakusta and Landau (2012; Exp 2), and by nativist models of language acquisition like that proposed in Jackendoff (1990). The latter argue that linking rules between conceptual and syntactic structures (e.g., agents → subject roles; patients → object roles) are ready from birth to guide successful mapping of forms into concepts (and on into syntactic frames). Lakusta and Landau found partial support for this position: in some contexts, children show a more pervasive goal bias across motional contexts than adults. This position was formalized with respect to the developmental trajectories of *come* and *go* in H1.

I test this possibility in the context of the deictic motion verbs *come* and *go*, whose adult use contrasts sharply in goal bias (Lester, 2014). If H1 is correct, then these verbs should both begin with equally large goal biases, only eventually differentiating *go* as a goal bias and *come* as a non-goal bias verb. If linguistic input drives children's usage more so than innate conceptual biases, then children's use of *come* and *go* should be different from the start, and possibly follow different trajectories.

Results 1

I fitted a linear mixed-effects model with the goal bias index as dependent variable; age estimates (in days), lemma (*come* or *go*), and their interaction as fixed effects; and speaker ids as random effect (intercepts only). A hierarchical backwards elimination of factors revealed that the maximal model, consisting of the two-way interaction between age and lemma, plus the two lower level main effects, could not be reduced any further (age:lemma: $F = 12.0396$, $p<0.001$). Closer inspection revealed a significant main effect for age ($F = 23.5018$, $p<0.001$) but not lemma ($F = 0.9828$, $p = 0.322$), indicating that the influence of word type on the goal bias index only becomes apparent when considered longitudinally. The significant two-way interaction is illustrated above as Figure 2.

Figure 2 plots the goal bias index (in characters) on the y-axis and the age spectrum (in days) attested in the Manchester corpus on the x-axis. The left panel shows the developmental trajectory of *come*; the right panel shows the same for *go*. The left panel shows no change in g, with sources and goals maintaining steady, close competition. The right panel shows a relatively stable increase in goal bias as the children age, with goals outweighing sources early, and with this difference increasing over time.

Discussion 1

This pattern of results does not support H1: the goal bias is clearly not the sole determinant of path description in children's early use of *come* and *go*. Rather, the fact that path encoding is treated differently for these verbs provides support for Lakusta and Landau's (2012) suggestion that presuppositions might drive patterns of mentioning. As Lester (2014) found for adults, children's path mentioning for *come* and *go* appears to be determined by their complementary path-based presuppositions: *come* presupposes goals and so occurs with them less frequently than sources; the opposite is true for *go*. Additionally, the persistence of the divide across the developmental span studied here suggests that children acquire this lexical contrast quite early, hinting at an input-driven effect. It may be that children simply encode the same types of paths that they have heard encoded with particular verbs, and that

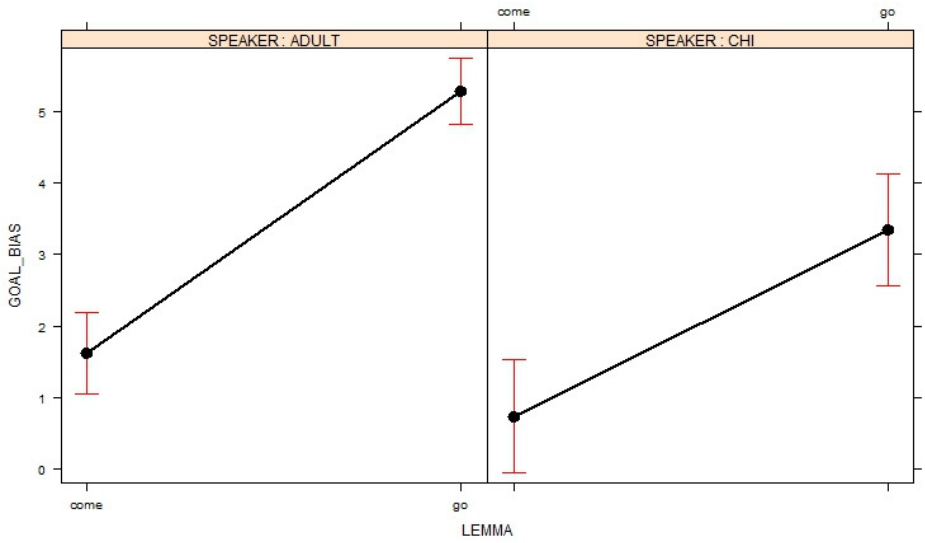

Figure 3: Goal bias as a function of the interaction of speaker type (*chi(ld), adult*) and verb (*come, go*)

children are sensitive to differences in the distributions of path-mentioning independent of the goal bias.

Turning to the individual verbs, the pattern for *go* (Fig. 1, right panel) could arise for several reasons. First, it might be that the proportional relationship between goal and source landmark encoding is maintained (or at least not reversed) during the general developmental trend towards longer path descriptions overall. A positive slope under these circumstances would indicate (a) that goals were consistently longer than sources by a particular margin (e.g., 'goals are always 2.5 times longer than sources') or (b) that either goals or sources grow somewhat faster, but that this rate is fixed. However, there are good reasons to suspect that neither of these is the case. Chief among these is the fact that children simply do not produce multiple path components in a single clause. Therefore, the means reported here do not reflect competition within the clause, but asymmetries in the complexity of mentioning of isolated sources and goals. Goal landmarks tend to get longer (perhaps due to general expansion of NP complexity) while sources continue to be omitted. *Go* therefore exhibits a standard goal bias, as expected based on Stefanowitsch and Rohde (2004) and Lester (2014).

The flat, low developmental slope for *come* suggests two things. First, *come* does not exhibit much of a goal bias ever. Second, this lack of a goal bias is not influenced by the children's increasing ability to produce more complex path descriptions. That is, neither goals nor sources offer themselves as loci for the development of landmark encoding in *come* clauses. Children's path-mentioning in *come* clauses is thus completely resilient to any influence of goal bias in general and in developmental terms.

Study 2: Path-mentioning in child and adult speech

The data from Study 1 suggest that the goal bias is strongly lexically mediated, and that the strength of the goal bias can change over time (contra H1). The question remains, however, to what extent this lexical mediation is a direct reflection of adult patterns of use (H2, above). To investigate this question, I added the adult data to the child data from Study 1, collapsed age into a binary class (adult vs. child) and performed a similar analysis. If adults and children behave similarly, this will be evidence that statistical biases in linguistic input can override conceptual biases.

Results 2

I fitted a second linear mixed-effects model with the goal bias index as dependent variable; speaker (*child, adult*) and lemma as fixed effects; and speaker IDs as random effect (intercepts only). A hierarchical backwards selection procedure revealed that the maximal model, consisting of the two-way interaction between speaker and lemma, plus the two lower level main effects, could not be reduced any further (lemma:speaker: $F = 12.55$, $p<0.001$). Closer inspection revealed a significant main effect for speaker ($F = 9.51$, $p<0.01$) and lemma ($F = 446.39$, $p < 0.001$). The significant interaction is illustrated as Figure 3 above.

Figure 3 plots the mean goal bias scores for *come* and *go* for adults (left panel) and children (right panel). Both adults and children use longer goal descriptions relative to source descriptions in *go* clauses as opposed to *come* clauses (accounting for the significant main effect of lemma). However, children showed reliably lower goal bias scores compared to adults for both verbs (main effect of speaker), with a more drastic difference arising for *go* than *come* (interaction of speaker and lemma).

Discussion 2

H2 was supported by these findings: in general, the patterns of usage in adult speech match the child-language data better than the universal-goal-bias position. The fact that an interaction between lemma and speaker did arise is likely

due to a general limitation on children's ability to produce path descriptions (that is, the higher *go* mean in the right panel of Figure 2 may be approaching the goal bias ceiling for children). Otherwise, the children reflect the same lexical contrast as the adults, *even those contrary to the goal bias*. Moreover, while we can attribute the adult pattern to Gricean strategies governing presupposition and omission, children at this age have not yet mastered these abilities. This suggests that the input (situated experience of language use) drives the acquisition of *come* and *go* above and beyond the underlying conceptual bias.

General Discussion

The results reported here have shown that the goal bias can be overridden by principled differences in the use of verbs otherwise fully grammatically licensed to occur with both goal and source adjuncts. Moreover, they demonstrated that this overriding effect of the linguistic input was present from the earliest ages (around 20 months).

This study is the first to confirm the presence of a goal bias in the naturalistic conversational speech of children, as well as the first to consider longitudinal data for individual children. In this way, I expand on the between-group analyses of Lakusta & Landau (2012). Adopting a longitudinal perspective revealed an interaction between goal bias and lexical semantics such that each motion deictic verb develops a distinctive path-structural representation.

Here, as predicted in Lakusta and Landau (2012), lexical variation in path-mentioning is sensitive to pragmatic presupposition. The contrasts in the presuppositional structure of deictic paths in *come* and *go* lead to statistical biases in the linguistic input, which become lexicalized components of the verbs' meanings over the course of development. Moreover, these linguistic biases are fully capable of overriding any 'hardwired' conceptual biases. The clear role of naturalistic input suggests that the effects observed by Lakusta and Landau could also receive an input-driven explanation (as opposed to invoking innately specified architecture for co-mapping forms across different semantic domains). Below, I introduce some possibilities for exploring this question further.

Future directions

This discussion opens up a variety of avenues for further exploration of goal bias and language acquisition. A first step might be to apply the methodologies used here to a wider array of verb types, including especially manner-of-motion verbs. In this way, one could determine whether the goal-inhibiting effects found for *come* are a fluke of deictic motion, or whether usage is the primary determinant for other verbs, as well. Also, cross-linguistic comparison of verb and path types both in adult and child speech would help to address questions of innateness vs. social/statistical-learning.

More psycholinguistic data is necessary to understand the verb-specific expectations about path component 'attentional worthiness.' One way to get at this issue would be to use an eyetracking paradigm. For instance, one could investigate preferential looking patterns within an array of landmarks while participants (adults and young children) are presented with recordings of various motion event descriptions. These studies would address the on-line comprehension of linguistic motion events relative to visual stimuli, and so present a more direct picture of the task faced by children acquiring these expressions.

References

Clark, E. V. & Garnica, O. K. (1974). Is he coming of going? On the acquisition of motion deictic verbs. *Journal of Verbal Learning and Verbal Behavior*, *13*, 559-572.

Jackendoff, R. (1990). *Semantic structures*. Cambridge: MIT Press.

Lakusta, L. M. (2005). Source and goal asymmetry in non-linguistic motion event representations (Unpublished doctoral dissertation). The Johns Hopkins University, Baltimore.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*, 1-33.

Lakusta, L. & Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive Science*, *36*(3), 517-544.

Lakusta, L, Wagner, L, O'hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language learning and Development*, *3*(3), 179-197.

Lester, N. (2014). The comings and goings of *come* and *go (and move)*. Paper presented at the American Association for Corpus Linguistics (AACL) 2014, Flagstaff.

Li, H. & Fang, A. C. (2011). Word frequency of the CHILDES corpus: Another perspective of child language features. *ICAME*, *35*, 95-116.

Papafragou, A. (2010). Source-Goal Asymmetries in Motion Representation: Implications for Language Production and Comprehension. *Cognitive Science*, *34*, 1064-1092.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.

Stefanowitsch, A. & Rohde, A. (2004). The goal bias in the encoding of motion events. In G. Radden & K. Panther, *Studies in linguistic motivation* (249-268). New York: Mouton de Gruyter.

Talmy, L. (2000). *Toward a cognitive semantics*, *Volumes 1/2*. MIT Press: Cambridge.

Theakston, A. L., Lieven, E. V. M., 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*, 127-152.