Syntactic Bootstrapping and the Internal Structure of Causative Events

Ann Bunger and Jeffrey Lidz
Northwestern University

1. Introduction

It has been well established that there exist systematic relationships between the way that we categorize events and the way that we describe them, and that these relationships have considerable effects on language processing and acquisition. The theory of syntactic bootstrapping, in particular, is based on the hypothesis that language learners can use their knowledge of these kinds of regularities to make conjectures about the meanings of new verbs. In support of this theory, numerous studies have demonstrated that children can be guided in the acquisition of novel verbs by precisely those aspects of syntax that are predicted by event structure representations (e.g., Landau & Gleitman 1985, Gleitman 1990, Fisher 1996, Lidz et al. 2003). What is not yet clear, however, is the scope to which this ability may be usefully applied. While it has been demonstrated that children can make use of syntactic bootstrapping to determine which of two distinct but simultaneous events is being labeled (e.g., Naigles 1990), it is unknown whether this ability can also be used to tease apart the subparts of a single complex event, or even whether children represent events in the same way that adults do.

Within the cognitive sciences, two significant avenues of research have been followed in the investigation of the nature of event representations. One of these is comprised of attempts to study event representations directly through experiments on event perception and categorization in the absence of language. From this research, we know that children can categorize events, some event types by as early as 2.5 months (Spelke et al. 1992, Baillargeon 1998), and we also know that this categorization is based on children's sensitivity to different combinations of event features, e.g., the identity and number of event participants (Leslie 1982, 1984; Baillargeon 1998; Gordon 2003), contact between event participants (Leslie 1984, Baillargeon 1998), and spatiotemporal continuity of motion (Leslie 1984, Spelke et al. 1992).

The second avenue of research into event representations takes the way that we use language to talk about events as a clue to the details of event structure representation. Numerous researchers have observed that there are regular structural relationships between syntactic and conceptual representations of events, such that, for example, transitive verbs, which take (at least) two noun phrase arguments, describe events that involve associations between two (or more) participants. It has been argued, furthermore, that these relationships between verb meaning and verb syntax rely on fixed mapping rules (e.g., Levin 1993, Baker 1997). Levin (1993; building on the work of Gruber 1965, Carter 1976, Jackendoff 1990, etc.) provides support for the claim that the meaning of a word determines its syntactic behavior by demonstrating that verbs fall into distinct subclasses on the basis of shared components of meaning that constrain syntactic behavior.

One example of this kind of fixed relationship between a verb's meaning and its syntactic behavior can be seen in the set of verbs that can participate in the so-called causative/inchoative alternation, illustrated in (1). Note that the verb *bounce* can be used both in a transitive frame, as in (1a), and in an unaccusative intransitive frame, as in (1b), in which the object of the transitive sentence appears as the subject of the intransitive.

(1) a. The girl bounced the ball.
    b. The ball bounced.

Verbs that can participate in this alternation share certain elements of meaning. First, these verbs must describe events that are internally complex: here, the girl performs some action, and this action causes a change of state in the ball. Crucially, furthermore, verbs that can occur in this alternation must label the object's change of state: it's the ball that bounces, not the girl. Other verbs that can participate in this alternation include *spin* and *roll*.

The verb *hit*, however, cannot participate in this alternation:

(2) a. Sammy Sosa hit the ball.
    b. * The ball hit.

Although *hit* can be used in a transitive sentence, as in (2a), the corresponding unaccusative intransitive sentence is ungrammatical (2b). Note that an event of hitting does not entail that the thing hit undergoes a change of state: this is why we can make observations like that in (3):

(3) John hit the tree trunk, but it wasn't affected.

Thus, the ungrammaticality of (2b) is due to the fact that the verb *hit* labels only an activity and not the result of that activity: it's Sammy that hits, not the ball.

Given that there exist these strict mappings between verb meanings and verb syntax, it seems logical to postulate that if these generalizations are part of

the tools that children have for language learning, they should be able to use these kinds of regularities to break into the linguistic system: that is, if they know the meaning of a verb, learners should be able to use that information to figure out something about the syntax of constructions in which it occurs, and on the flip side, if they know something about the syntactic frame in which a verb occurs, they should be able to use that information to figure out something about the verb's meaning.

The syntactic bootstrapping hypothesis (Landau & Gleitman 1985, Gleitman 1990, Gillette et al. 1999), which corresponds to the latter of these two possibilities, depends crucially on these mapping rules. The central claim of this theory is that children can observe the syntactic structures in which a novel verb occurs and then use that distributional information and the mapping rules (coupled with observation of the real-world context of use) to infer important information about the meaning of the verb: if a verb is used in X way, then it must be describing Y event or Y perspective on some event.

Studies carried out by Naigles and colleagues (Naigles 1990, Naigles & Kako 1993, Naigles 1996) have demonstrated that the use of syntax in verb learning begins very early: by the age of 24 months, children may use information from the syntactic frame (transitive vs. intransitive) in which a novel verb is presented to determine the meaning of that verb. Naigles (1990) demonstrates that, given a scene depicting two simultaneous events (a causative event and a non-causative continuous event), the structure of the sentence in which a novel verb is presented can influence a child's interpretation of the meaning of that verb. Specifically, children who heard a novel verb in a transitive sentence (*X is gorping Y*) interpreted it as labeling the causative action, whereas those who heard the novel verb in an intransitive sentence (*X and Y are gorping*) interpreted it as labeling the continuous action. Naigles (1996) demonstrates, furthermore, that children can use multiple syntactic frames to help determine the meaning of a novel verb.

There is compelling evidence, then, in support of the proposal that syntactic bootstrapping allows the child to use clues from syntax learn verb meanings. It is important to keep in mind, moreover, that the meanings that children assign to new verbs are presumably informative about their representations of the events labeled by those verbs. Given that fact, we should be able to turn the tables and use the way that children use syntax to interpret an event to learn something about the things they represent as part of that event: i.e., we can use the systematic relationship between syntax and verb meaning as a clue to what event representations look like. The goal of the research reported here is to use children's sensitivity to this regular mapping between verb syntax and meaning to probe how they represent a single event that is internally complex: i.e., to see what happens when a child is presented with a scene depicting a single event that might be divided into several different subevents (e.g., a cause and an effect) and is asked to determine which of these subevents is being labeled on the basis of the syntactic information provided by the input sentence.

Conveniently, causative verbs (e.g., *kill*, *roll*) provide just the right kind of event structure to investigate this question. Verbs of this subclass, although lexically one unit, describe events that can be decomposed into two subevents: the thing the agent is doing and the resultant state or motion of the patient or affected object. It has been suggested, furthermore, that at some level the linguistic representation of lexical causatives reflects this internal structure (e.g., Jackendoff 1990, Levin & Rappaport Hovav 1995, Harley 1996). The conceptual structure of a causative event might be represented as in (4), with open argument positions to be filled by the causer of the event and the affected entity:

(4) [[X do something] CAUSE [Y become state]]

This first subpart of this structure [X do something] specifies the causing subevent, or the means, and the second subpart [Y become state] specifies the resulting change of state.

There are linguistic reasons to think that we represent these events with this kind of internally complex structure. Note that we can provide linguistic descriptions of the various subparts of a causative event representation. We can refer to an event in which a girl bounces a ball, for example, by specifying just the activity of the girl, as in (5a), or we can specify only the resulting change of state in the ball, as in (5b), or we can describe the entire event, as in (5c).

(5) a. The girl is patting the ball.
     b. The ball is bouncing.
     c. The girl is bouncing the ball.

At issue for the study of language acquisition is whether young children have access to the same complex representations for these events that adults do, and, if so, whether syntax can direct their attention to the subevents (thereby influencing their interpretation of a novel verb used to describe the event). To think about this from the other direction, if syntax can direct a learner's attention to the subevents of a complex event, then this is evidence that that learner's event representations are internally complex.

In terms of syntactic bootstrapping, furthermore, the question is whether children are limited to using this ability to distinguish between multiple distinct events in the world, or whether they can also use it to parse single events that are internally complex. Recall that the studies by Naigles and colleagues demonstrate that when presented with two simultaneous events, 2-year-olds can use their knowledge of the regularities between syntax and semantics to choose which of those events is being referred to by a novel verb. In our experiments, children are presented instead with a single event that is internally complex, like these causatives, to find out whether, when there's only one event going on, syntax can shine a spotlight on one or the other of the subparts of the event to help children decide what a novel verb is referring to.

In this study, children are presented, in a preferential looking paradigm, with causative events described either with a sentence including no novel verb or with a nonsense verb label given in a transitive frame, an unaccusative intransitive frame, or a combination of the two and are asked to decide which subevent of the complex causative that verb labels: the means or the result. If children are aware of the multiple ways that this complex event structure can be interpreted (i.e., they represent the subevents that make up a causative event), the meaning that they assign to a novel verb describing such an event should change with the syntactic frame in which it is presented. More specifically, because the unaccusative variant of a causative verb labels the result subevent without making reference to the means, subjects in the unaccusative and multiple frame conditions should be more likely than subjects in the transitive and control conditions to interpret the novel verb as referring to the result subevent.

2. Methods

This experiment employs the preferential-looking paradigm developed by Spelke (1979) and Golinkoff et al. (1987) to study intermodal perception in infants. In this procedure, a child is presented with two scenes displayed simultaneously on opposite sides of the screen of a large video monitor accompanied by some speech stimulus. Previous studies (e.g., Golinkoff et al. 1987) have shown that children tend to look longer at the scene that matches the speech stimulus.

2.1 Subjects

The subjects were 44 children (11 in each experimental condition) ranging in age from 22;1 (months;days) to 25;29 (mean 24;10). All were being raised in English-speaking homes. An additional 14 children were run in the experiment, but were excluded from analysis for one of the following reasons: age, language background, unwillingness to complete the experiment, fixation bias of more than 75% to a single side of the video monitor, or inattention during the test phase for more than 30% on two or more trials.

The period of development around 2 years of age, which corresponds approximately to the two-word stage of performance, is crucial for studies of language acquisition because the child still has a lot of words to learn and yet is just beginning to use syntax in her own production. Logically, then, it is at this time that a mechanism like syntactic bootstrapping would be most useful for the child, and, interestingly, it is exactly during this stage that the child experiences a rapid increase in the acquisition of new vocabulary, and especially of verbs.

2.2 Stimuli

The stimuli consisted of color videos depicting four causative events for presentation during the training phase of each trial and eight subevents (two per causative event) for presentation during the test phase. A complete list of the events and subevents used as stimuli is given in Table 1: all were performed by live actors and were recorded against a neutral background in a bare environment. In a given training phase, subjects were presented with a causative event accompanied by a description of the event incorporating a novel verb. Causative events were presented on both sides of the video monitor, first on each side in sequence for 6 seconds each, and then on both sides simultaneously for another 6 seconds.

Table 1—Video events

| Novel verb | Causative event | Subevents | |
		Means	**Result**
lorp	boy pulls wagon	boy tugs on wagon that doesn't move	wagon rolls; boy does nothing
pim	girl bounces ball	girl pats ball that doesn't move	ball bounces; girl does nothing
flurb	girl pushes chair across the floor	girl pushes chair that doesn't move	chair rolls; girl does nothing
blick	boy spins garden flower	boy waves hand in front of flower that doesn't move	flower spins; boy does nothing

Experimental conditions differed between subjects and were distinguished by the kind of auditory input subjects received during training. Subjects in the control, or no word, condition, were presented with video of a causative event accompanied by an auditory stimulus that invited them to attend generally to the event. Subjects in the unaccusative condition saw the causative event accompanied by a novel verb presented in an unaccusative intransitive sentence; subjects in the transitive condition heard the novel verb presented in a transitive sentence; and subjects in the multiple frame condition heard the novel verb used in both of these syntactic frames—first in a transitive sentence, and then in an intransitive sentence. All auditory stimuli were recordings of a female talker using child-directed speech intonation and were added to the audio track of the videos in synchronization with the video output. For the unaccusative, transitive,

and multiple frame conditions, each presentation of the causative event was accompanied by two uses of the novel verb (for a total of six presentations of the novel verb). Examples of the auditory stimuli are provided in Table 2.

Table 2—Experimental conditions

Syntactic frame	Example of auditory input
No word	"Wow—look at that! What's happening here?"
Unaccusative	"Hey look! The ball is pimming. Do you see the ball pimming?"
Transitive	"Hey look! The girl is pimming the ball. Do you see the girl pimming the ball?"
Multiple frame	"Hey look! The girl is pimming the ball. Do you see the ball pimming?"

On each trial, this training phase was followed by a test phase in which subjects were presented with the two subevents related to the causative event presented in the training phase (i.e., the means and result subevents). These two subevents were presented simultaneously on opposite sides of the screen for a total of 12 seconds, and the auditory stimulus directed the child to find the action represented by the novel verb introduced during training. A schematic representation of the video scenes and auditory accompaniment (for the unaccusative condition) is presented in Table 3.

The side of the video monitor on which the causative event was first presented during the training phase was counterbalanced, as was the location (right vs. left side) of the means and result subevents shown during the test phase.

2.3 Procedure

Children were tested individually, seated on a booster chair facing the video monitor. Parents stood just behind the chair and were asked to wear a visor over their eyes while in the testing room to limit their exposure to the stimuli.

The child's attention was centered before each trial by the presentation of a 5-second (re)centering stimulus (video of a giggling baby), and then during the trials, children were free to sample the contents of both sides of the video

monitor. The child's attention to the stimuli was recorded using a digital video camera situated on top of the video monitor.

Table 3—Schematic representation of experimental design

Phase	Left side of screen	Audio track	Right side of screen
Training	girl bouncing ball	Look! The ball is pimming. Do you see the ball pimming?	black screen
	black screen	Hey look! The ball is pimming. Do you see the ball pimming?	girl bouncing ball
	girl bouncing ball	Wow! The ball is pimming. Do you see the ball pimming?	girl bouncing ball
Test	girl patting immobile ball	Oh look! They're different. Do you see pimming?	ball bouncing on its own
	girl patting immobile ball	Do you see pimming? Where's pimming now?	ball bouncing on its own

2.4 Analysis

Videos made of the subjects were coded for direction of visual fixation (left vs. right) during each frame of the test phases (360 frames per trial, 4 trials per subject). An ANOVA was performed to test for differences in mean visual fixation to the means vs. result subevents across experimental conditions, and paired t-tests for differences within conditions. All P values are two-sided and are significant at the 0.05 level.

3. Results

The results of previous studies suggest that the meaning that children assign to a novel verb describing some event should change with the syntactic frame in which that verb was presented. Assuming that the child assigned a meaning to the novel verb during the training phase in this experiment, one of the subevent scenes should provide a better match for the auditory stimulus presented at test, with that match crucially dependent on how the child interpreted the verb. The relevant question to ask when examining these data, then, is whether differences

in the frame in which novel verbs were presented during the training phase translated into differences in the attention that subjects paid to the two subevents presented during the test phase.

Figure 1 depicts the data on direction of visual fixation at test for each experimental condition, averaged across subject and trial. ANOVA testing revealed a significant main effect of subevent preference ($p<0.0001$). Across conditions, attention to the result subevent was slightly greater than that to the means subevent during the 12-second test phase.[1] There was also a significant interaction between condition and subevent preference ($p=0.0015$): preference for the result subevent was significantly greater for the unaccusative (F=20.65, $p=0.0011$) and multiple frame (F=9.15, $p=0.0115$) conditions than for the other two conditions (no word: F=1.8638, $p=0.2021$; transitive: F=4.0861, $p=0.0708$).
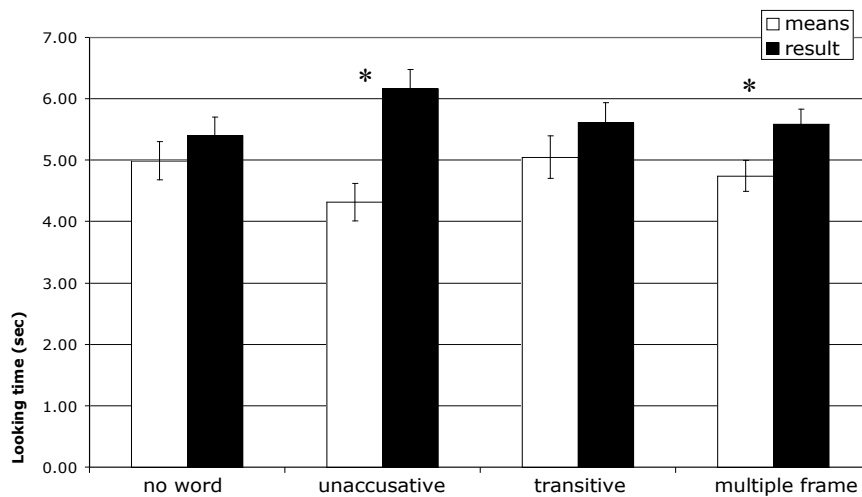

Figure 1—Mean visual fixation at test
* Difference in attention to the means vs. result subevent was significant in the unaccusative and multiple frame conditions.

Taking attention to subevents in the no word condition as an indicator of baseline preferences, the data show that there is a slight but insignificant trend toward a baseline preference for the result subevent. In the unaccusative and

1. This overall preference for the result subevent was not consistent across all trials. Note, however, that the trends observed in Fig. 1 are mirrored in the preferences demonstrated by condition in each event: for each event, subjects demonstrated a greater preference for the result subevent in the unaccusative and multiple frame conditions compared with the no word and transitive conditions.

multiple frame conditions, this natural preference is enhanced: the trend toward a preference for the result subevent is significant. In the transitive frame condition, however, subjects do not demonstrate a preference that is significantly different from chance.

4. Discussion

One of the tools that children use in the acquisition of novel verbs is their knowledge of the consistent projection of linguistic structure from conceptual representations. As detailed in the introduction, previous studies (Gleitman 1990, Naigles 1990, etc.) have demonstrated that children can use clues from syntax to help them decide which of a pair of events a novel verb refers to. What the results of this study demonstrate is that children can also use clues from syntax to determine which perspective on a single event is being described.

In this study, children exposed to the same causative event interpreted a novel verb associated with that event differently depending on the syntactic frame in which that verb was presented. In the absence of syntactic clues, i.e., in the no word condition, subjects showed a slight preference for the subevent that corresponded to the result of the causative event. Presentation of a novel verb in an unaccusative intransitive frame enhanced this preference, whereas presentation of the same novel verb in only a transitive frame did not. This pattern of results can be explained by taking into account the requirements on verbs occurring in these syntactic frames when used to label a causative event.

Recall from the discussion of the mapping between verb meaning and verb syntax above (see example (1), repeated here for convenience as (6)) that any verb that can participate in the causative alternation must be labeling the result of that event.

(6) a. The girl is bouncing the ball.
     b. The ball is bouncing.

To present this phenomenon from another angle, it can be said that verbs labeling (agent-induced) causative events that can appear in an unaccusative intransitive syntactic frame must be labeling the result. Thus, because subjects in the unaccusative condition in this experiment were presented with an event description that drew their attention to the result of the causative (and that left out the means altogether), they interpreted the novel verb as a label for that subevent.

Verbs presented in a transitive frame, on the other hand, are trickier to decode. As illustrated in (5) above (repeated here in part as (7)), the verb in a transitive sentence used to describe a causative event can label either the means of the event (7a) or the result (7b).

(7) a. The girl is patting the ball.
     b. The girl is bouncing the ball.

Thus, because subjects in the transitive condition in this experiment received no unambiguous clue as to which subevent the novel verb is labeling, their interpretation of its meaning was at chance. By the same token, because the only unambiguous presentation of the novel verb that subjects in the multiple frame condition in this experiment were exposed to was as a label for the result (i.e., in an unaccusative intransitive sentence), they interpreted it as referring to that subevent.

This study provides further evidence, then, that observation of the syntactic behavior of a novel verb provides information about the kind of event that the verb labels. Note that if the syntactic nature of the input did not influence the interpretations that subjects assigned to novel verbs in this study, then we should not have observed this difference in attention at test between the unaccusative and multiple frame conditions on the one hand and the no word and transitive conditions on the other.[2] Likewise, if the subjects were not representing these causative events as having internal structure, the syntax would not have been able to guide them to these different interpretations: i.e., if their representations of these causative events did not include a result subpart, then even subjects in the unaccusative and multiple frame conditions would not have been able to tease that subevent apart from the whole event.

5. Conclusions

The results of this study demonstrate that 2-year-old children do represent complex causative events as having internal structure, and moreover, that they can use syntax to identify and label the subparts of internally complex events. Here, when the linguistic input focused children's attention on the result of a given causative event, they interpreted a novel verb as describing that subevent. More generally, then, it seems that we can, in fact, use syntactic bootstrapping not only to tell us about the way that labels for events are acquired, but also to tell us something about how those events are represented by the learner.

---

2. Data collected for an additional condition in this experiment argue against the potential criticism that the results reported here are due to children's sensitivity to clues provided by co-occurring nouns in the input rather than to clues provided by syntax. In this condition, novel verbs labeling these causative events were presented in an unergative intransitive frame, as in (i):

(i) The girl is pimming.

If 2-year-olds are only paying attention to co-occurring nouns, we would expect them in this condition to interpret the novel verb as a label for the means subevent, since it is the only subevent in which the agent plays an active role. On the contrary, however, the data indicate that they are even more likely to interpret the verb in an unergative intransitive frame as a label for the result.

References

Baillargeon, R. (1998). Infants' understanding of the physical world. In Sabourin, M. & Craik, F. (eds.), *Advances in Psychological Science, Vol. 2: Biological and Cognitive Aspects*. Hove, England: Psychology Press/Erlbaum.

Baker, M.C. (1997). Thematic roles and syntactic structure. In Haegeman, L. (ed.), *Elements of Grammar: Handbook in Generative Syntax*. Dordrecht, Netherlands: Kluwer.

Carter, R. (1976). Some constraints on possible words. *Semantikos*, 1, 27–66.

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

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

Golinkoff, R.M., Hirsh-Pasek, K., Cauley, K.M., & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14, 23–45.

Gordon, P. (2003). The origins of argument structure in infant event representations. 28th Annual Boston University Conference on Language Development, Boston, MA, October 31–November 2, 2003.

Gruber, J. (1965). *Studies in Lexical Relations*. Dissertation, MIT.

Harley, H. (1996). Events, agents and the interpretation of VP-shells. Unpublished ms. University of Pennsylvania.

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

Landau, B. & Gleitman, L.R. (1985). *Language and Experience: Evidence from the Blind Child*. Cambridge, MA: Harvard UP.

Leslie, A.M. (1982). The perception of causality in infants. *Perception*, 11, 173–186.

Leslie, A.M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, 13, 287–305.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago.

Levin, B. & Rappaport Hovav, M. (1995). *Unaccusativity at the Syntax–Lexical Semantics Interface*. Cambridge, MA: MIT Press.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: verb-learning and the footprint of universal grammar. *Cognition*, 87, 151–178.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17, 357–374.

Naigles, L.R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58, 221–251.

Naigles, L. & Kako, E.T. (1993). First contact in verb acquisition: defining a role for syntax. *Child Development*, 64, 1665–1687.

Spelke, E.S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, 15, 626–636.

Spelke, E.S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99, 605–632.