How Children and Adults Encode Causative Events Cross-Linguistically: Implications for Language Production and Attention

4 authors, including:

Ann Bunger
University of Delaware
16 PUBLICATIONS   **115** CITATIONS

SEE PROFILE

Dimitrios Skordos
The University of Calgary
12 PUBLICATIONS   **65** CITATIONS

SEE PROFILE

Anna Papafragou
University of Delaware
112 PUBLICATIONS   **2,690** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Beyond words: children apply the principle of informativeness to non-linguistic symbols. View project

Neural correlates of scalar implicature derivation in preschool children and adults View project

How Children and Adults Encode Causative Events Cross-Linguistically:

Implications for Language Production and Attention

Ann Bunger[a,*,1], Dimitrios Skordos[a,2], John C. Trueswell[b], Anna Papafragou[a,*]

[a] Department of Psychology, University of Delaware, 108 Wolf Hall, Newark, DE 19716 USA

[b] Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104 USA

acbunger@indiana.edu

dskordos@udel.edu

trueswel@psych.upenn.edu

papafragou@psych.udel.edu

* Corresponding authors. Tel.: +1 302 831 2752, Fax: +1 302 831 3645

Word count: 12244

[1] Present address: Department of Linguistics, Indiana University, 322 Memorial Hall, 1021 E. Third St., Bloomington, IN 47405 USA
[2] Present address: Department of Psychology, 3535 Mandler Hall, University of California San Diego, La Jolla, CA 92093 USA

Abstract

This study investigates the implications of language-specific constraints on linguistic event encoding for the description and on-line inspection of causative events. English-speaking and Greek-speaking adults, 3-year-olds, and 4-year-olds viewed and described causative events, which are composed of Means and Result subevents, in an eyetracking study. The results demonstrate cross-linguistic differences in the informational content of causative event descriptions: Greek speakers across age groups were more likely than English speakers to mention only one causative subevent. Developmental changes in the tendency to encode information about causative events in language were also evident: in both language groups, adults were more likely than children to mention both Means and Result subevents. Finally, for both adult and child speakers of both languages, preparing different types of event descriptions changed the way that events were visually inspected, shifting attention toward to-be-encoded subevents. These findings offer some of the first evidence about the development of the language production system, the attentional mechanisms that it employs, and its workings in speakers of different languages.

  Keywords: causative events, cross-linguistic differences, language production, event cognition

How Children and Adults Encode Causative Events Cross-Linguistically:

Implications for Language Production and Attention

The world around us is a complex and dynamic place, with a constantly varying array of

people and objects moving into and out of a variety of different relationships with each other. In

order to make sense of this dynamic set of stimuli, the human mind creates structured

representations of meaningful chunks of activity by pulling together abstract spatial, temporal,

and causal information about the world—what we refer to as *event representations*. When we

describe events, we have to make choices about what subset of information about a given event

we want to convey and how to encode that information in language. As is well known, however,

languages differ both in the kinds of information they select when encoding events and in the

way this information is packaged into syntactic-semantic structures (e.g., Talmy, 1985, 2003).

Clearly, such cross-linguistic variation needs to constrain theories of how humans map event

representations onto language and of how this process might change over development.

Much, perhaps most, of the work on the cross-linguistic encoding of events in adults and

children has focused on the domain of spontaneous motion (e.g., Allen et al., 2007; Kersten et

al., 2010; Malt et al., 2008; Naigles, Eisenberg, Kako, Highter, & McGraw, 1998; Özyürek et al.,

2008; Papafragou, Massey, & Gleitman, 2006; Slobin, 2004). This domain is characterized by

considerable cross-linguistic variation. For instance, one group of languages including English,

German, and Swedish, typically encode manner of motion information in the verb ("He ran…")

and path information in a modifier ("into the store"), while a second group of languages,

including Greek, Spanish, and Turkish typically encode path information in the verb (e.g., Greek

"Bike…," 'enter-3sing-past'), perhaps with additional path modifiers ("sto magazi," 'in(to) the

store'), and might omit manner information or encode manner in a modifier ("trehontas,"

'running'). Several studies have shown that both children and adults are sensitive to language-

specific patterns of motion event description, and these patterns emerge as early as 3 years of age

(Allen et al., 2007; Naigles et al., 1998; Özyürek et al., 2008; Papafragou et al., 2006; Slobin,

1996; Zheng & Goldin-Meadow, 2002). Furthermore, language-specific motion encoding

preferences lead both children and adults to make systematically different (path- vs. manner-

oriented) conjectures about what novel motion verbs might mean in their native language

(Maguire et al., 2010; Naigles & Terrazas, 1998; Papafragou & Selimis, 2010; Skordos &

Papafragou, 2014).

Particularly relevant for present purposes is evidence that, for adults, cross-linguistic

differences in motion encoding are reflected in differences in patterns of event inspection during

speech planning. Papafragou and colleagues (2008) demonstrated that the attention that adult

speakers of English and Greek directed to motion events during language production varied with

cross-linguistic differences in the way the events were described, with both groups directing

more attention to aspects of the scene that they planned to encode in a verb (i.e., Greek speakers

to paths and English speakers to manners of motion). Importantly, the overt attention directed to

the manner and path of motion was identical across the two language groups during free-viewing

of the same motion stimuli (see also Trueswell & Papafragou, 2010; cf. Bunger, Papafragou &

Trueswell, 2012 for extensions of this work to English-speaking 4-year-olds). These results

confirm and extend single-language demonstrations showing that, when planning speech, people

very quickly direct their gaze to components of the scene that they plan to talk about, usually in

the order that they plan to speak about them (Bock, Irwin, Davidson, & Levelt, 2003; Gleitman,

January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000; Griffin & Spieler, 2006). They also

offer the first demonstration that the online mobilization of linguistic resources for purposes of communication—often referred to as "thinking for speaking" (Slobin, 1996; cf. Levelt, 1989)—affects attention differently in adults who speak different languages.

Despite the wealth of evidence summarized so far, the cross-linguistic study of event encoding currently faces several limitations. First, the emphasis on the domain of spontaneous motion severely restricts the empirical foundations of theories of how language and event perception make contact with each other across languages and age groups. It is important to document cross-linguistic similarities and differences in other domains, including more complex events (see Evans, Gaby, Levinson & Majid, 2011; Kopecka & Narasimhan, 2012; Majid, Boster & Bowerman, 2008), and to track when and how language-specific patterns of event description are acquired across a broader sample of event types.

Second, cross-linguistic investigations of event encoding in both adult and child populations are just beginning to be combined with online measures of attention allocation as people produce event descriptions (e.g., MacDonald, 2013; Norcliffe, Konopka, Brown & Levinson, 2015; Papafragou et al., 2008). As a result, we lack a detailed understanding of how the language production system performs the meaning-to-form mappings necessary to successfully encode events within different languages. This gap in knowledge is particularly acute from a developmental perspective: even though language-specific event encoding has been documented as early as the age of 3 (e.g., Maguire et al., 2010; Skordos & Papafragou, 2014, on motion), little is known about the architecture of the production system in such young children (see McDaniel, McKee & Garrett, 2010). Furthermore, it is an open question whether speech planning directs attentional resources in children in the same ways that it does in adults. The cognitive mechanisms that regulate attentional control develop over time, and young children are

more limited than adults in, for example, their ability to divide their attention between multiple locations in a visual array (e.g., Scerif et al., 2005). It is thus important to establish whether, across languages, preparing to speak in adults and young children guides attention during event inspection, and whether these effects change between younger and older learners.

Currently there are several hypotheses about how cross-linguistic similarities and differences might shape language production (i.e., thinking for speaking) in both adults and children that remain mostly untested. At a broad level, one might expect that speakers across languages would attend to event components that they plan to mention in the order in which they plan to mention them regardless of native language (see Bock et al., 2003; Gleitman et al., 2007; Griffin & Bock, 2000; cf. also Levelt, 1989; Slobin, 1996). Thus (other things being equal) speakers who produce semantically and syntactically equivalent surface event descriptions across different languages should examine events in similar ways when preparing to describe them (e.g., see Norcliffe et al., 2015, for a comparison of formulation in adult speakers of Tzeltal and Dutch).

Beyond such broad effects, however, more specific relationships between typological variation and the process of language production remain open. For instance, speakers of two languages that produce similar surface linguistic descriptions of events might not have identical attention-allocation patterns prior to speaking if these formulations represent a strong typological preference in one language but a more peripheral pattern in the other. This idea is consistent with Slobin's (2004) discussion of effects of typological differences on production. Slobin suggested that if two languages encode a certain domain with some frequency, but the domain is relatively more codable in one language than the other, differences in codability should be reflected in thinking for speaking, with speakers of some languages "habituated to making frequent online

decisions" about the highly codable domain compared to others in which the domain is less codable and therefore "less salient" (p.175; cf. Levelt, 1989). This idea also coheres with proposals within the language production literature according to which speakers favor frequently used and at least partially lexically-independent abstract plans for building syntactic structures ("Plan Reuse," MacDonald, 2013): to the extent that the availability of such plans varies across languages, speech planning should be correspondingly affected. This possibility has not been explored in the literature to date. If sensitivity to the role of typologically preferred event encoding plans is reflected in the complexity or effort of speech planning, we would expect it to affect the way that speakers of different languages shift attention between to-be-encoded event components even when they produce similar linguistic outputs. Such effects on attention, if found, should be more pronounced for adults (who are more experienced language users) compared to young children.

Finally, for both adults and children, cross-linguistic studies of event inspection in communicative contexts need to be combined with explorations of event inspection in nonlinguistic contexts. This is because it is important to establish the extent to which eyegaze patterns observed during event description reflect the task of language production as opposed to the underlying salience of event subcomponents (see Bunger et al., 2012; Papafragou et al., 2008). At present, there are very few studies of nonlinguistic event perception, and even fewer that compare multiple age and/or language groups (see Shipley & Zacks, 2008, for a review). Furthermore, the relation of nonlinguistic event representations to language is a matter of debate, with some commentators expecting strong continuity of event representation across language communities and others expecting one's native language to affect nonlinguistic event representations themselves (for different perspectives see Griffin & Bock, 2000; Kersten et al.,

2010; Levinson, 1996; Papafragou et al., 2008, and the reviews in Gentner & Goldin-Meadow, 2003; Gleitman & Papafragou, in press; and Wolff & Holmes, 2011).

In this study, we address these open questions at the interface between language and events within a novel empirical domain, that of causative (caused motion) events. Causative events are of interest because they are complex events composed of multiple dynamic subcomponents that overlap in time and space. Several studies have looked at the way that infants, children, and adults reason about causative events (Bunger, 2006; Cohen, Rundell, Spellman & Cashon, 1999; Göksun, George, Hirsh-Pasek & Golinkoff, 2013; Leslie, 1984; Leslie & Keeble, 1987; Oakes & Cohen, 1990; Wolff, 2007, 2008) and the way that learners come to encode subcomponents of such events into different types of verbs (Behrend, 1990; Behrend, Harris & Cartwright, 1995; Bowerman, 1989; Bunger, 2006; Forbes & Farrar, 1993, 1995; Gentner, 1978; Gropen, Pinker, Hollander & Goldberg, 1991; Naigles, 1996; Papafragou & Selimis, 2010; Slobin, 1985). However, the cross-linguistic description of causative events has been the topic of only a few studies – and those mostly in adults (Klettke & Wolff, 2003; Wolff, Jeon, Klettke & Yu, 2010; Wolff, Jeon & Yu, 2009; Wolff & Ventura, 2009; but cf. Ji, Hendriks & Hickman, 2011).

Specifically, we ask how 3-year-old, 4-year-old, and adult speakers of English and Greek attend to causative events for the purpose of describing them. We focus on young children since, as mentioned already, this is the earliest age at which sensitivity to language-specific encoding patterns has been documented, and because children's event descriptions become increasingly rich throughout the preschool years (e.g., Allen et al., 2007; Özyürek et al., 2008; Papafragou et al., 2006; Slobin, 1996, on motion). We chose English and Greek because, as we discuss below,

these two languages place different constraints on the way that speakers encode information about causatives (see also Papafragou & Selimis, 2010).

Causative Events Across Languages

We focus on events of caused motion. In such events, an agent performs an action on an object which, in turn, causes the object to change location. Both children and adults conceive of such complex causal chains as being composed of two distinct subevents: a Means and a Result (e.g., Bunger, 2006; Cohen et al., 1999; Jackendoff, 1990; Leslie, 1984; Leslie & Keeble, 1987; Talmy, 1985). Imagine, for example, the sequence of activity in a soccer game in which a player kicks the ball, thereby sending the ball across the field and into the goal. The Means subevent corresponds to the action of the agent on the object (the player's kicking of the ball). The Result subevent corresponds to the change that the object undergoes (the movement of the ball to the goal). These interactions are interpreted as subevents of a complex causative event rather than as a sequence of unrelated happenings because of a perceived causal link between the Means and Result. That is, the viewer understands that the player's kicking of the ball directly resulted in the ball's change of location.

Even though cross-linguistically the encoding of causative events typically makes reference to Means and Result subevents (e.g., Jackendoff, 1990; Levin, 1993; Levin & Rappaport Hovav, 1995; Pustejovsky, 1991; Talmy, 1985), individual languages tend to differ in the ways they package information about causative events in semantic-syntactic structures (e.g., Talmy, 2003). English and Greek appear to be similar in the selective encoding of a single subevent within a causal chain – Means, as in (1), or Result, as in (2).

(1)   a. The player kicked the ball.

           b. O pektis klotsise ti bala.

              'the player kicked the ball'

(2)   a. The player sent the ball into the goal.

           b. O pektis estile ti bala sto terma.

              'the player sent the ball to the goal'

However, the two languages differ in the options available for speakers to encode

information about *both* causative subevents. English offers a standard structure for packaging

specific information about both causative subevents into a single sentence: English speakers can

combine a verb like "kick" that specifies a Means activity with a post-verbal prepositional phrase

like "into the goal" that specifies a Result, giving rise to the sentence in (3).

(3)   The player kicked the ball into the goal.

This packaging option is less available to Greek speakers than it is to English speakers:

there are fewer verbs in Greek that allow the addition of resultative phrases that specify the

Result or endpoint of a complex event (Giannakidou & Merchant, 1999; Papafragou et al., 2008;

Papafragou & Selimis, 2010). In fact, the Greek equivalent of the sentence in (3), given in (4), is

ungrammatical. Thus complex causative events are less codable in Greek than in English because

the compact and efficient event-packaging structure in (3) is more restricted (even when they are

grammatical, resultative phrases are less common in Greek as a whole). To encode the event in

(3), Greek speakers would have to choose between several options: (i) mention only one of the

two causative subevents, as in sentences (1b) and (2b); (ii) use a less specific means verb that allows a resultative phrase, as in (5a); or (iii) describe both subevents of this complex causative by stringing together multiple sentences, each of which encodes a single subevent, as in (5b), an option that has been linked to looser planning units in speech preparation (Özyürek et al., 2008).

(4)   * O pektis klotsise ti bala sto terma.

         'the player kicked the ball to the goal'

(5)   a. O pektis erikse ti bala sto terma.

         'the player threw the ball to the goal'

         b. O pektis klotsise ti bala ke afti pige sto terma.

         'the player kicked the ball and it went to the goal'

The Current Study

        The current study has two major goals. The first goal is to contribute to what is known about the cross-linguistic encoding of events by comparing the descriptions of caused motion events from 3-year-old, 4-year-old, and adult speakers of English and Greek. We begin by asking how cross-linguistic similarities and differences in information packaging affect the way that mature speakers of Greek and English describe causatives. Based on the typological differences we surveyed earlier, we might expect to find that adult speakers of English and Greek share the tendency to encode both causative subevents in their speech, but that English speakers are more likely to do so than Greek speakers. Furthermore, even if they include both subevents, the structure of speakers' descriptions should vary: in English, such descriptions should be more consistent, mostly conforming to the single-clause structure in (3), whereas in Greek,

descriptions should be more variable, including both one- and two-clause structures as in (5a) and (5b).

Relatedly, we ask how the description of causative events changes with development cross-linguistically, i.e., whether English- and Greek-speaking children describe causative events like adult speakers of their native language, and how early these patterns emerge. Recall that children are sensitive to language-specific patterns of event description for a simpler kind of event—motion events—by as early as 3 years of age (e.g., Maguire et al., 2010; Skordos & Papafragou, 2014). If language-specific patterns emerge early here as well, we might expect English- and Greek-speaking preschoolers to differ in the tendency to provide complete (Means+Result) descriptions of causative events, with children between 3 and 4 years becoming increasingly more similar to adults in their language community in terms of their event encoding patterns. Alternatively, or additionally, children from different language communities might be more similar to each other compared to adults in terms of their verbal descriptions of causatives. We know that children are more likely than adults to omit event information (cf. Allen et al., 2007; Bunger et al., 2012; Özyürek et al., 2008; Papafragou et al., 2006). It is an open possibility that 3- and 4-year-olds' causative descriptions will primarily encode a single causative subevent (cf. (1) and (2) in the previous section). If so, it would be of further interest to assess whether children across languages would converge or diverge in terms of the single subcomponent they choose to encode linguistically.

A second major goal is to examine the way that participants from different language backgrounds and age groups gather information about causative events in real time as they plan to speak. We are particularly interested in how cross-linguistic similarities and differences might impact the way speakers of English vs. Greek prepare to describe causatives. We first test the

expectation that verbal descriptions that include both causative subevents (arranged in the same order) would result in similar eye movement patterns regardless of language background. Beyond this global similarity, we test the further expectation that typologically preferred event encoding patterns shape frequent online decisions about how to package events into sentences (Slobin, 2004; cf. MacDonald, 2013). Specifically, we explore the possibility that English speakers' eye movements while preparing event descriptions might reveal a more streamlined process of attending to Means and Result in sequence compared to Greek speakers, even in cases in which both groups mention both subevents. That is, Greek speakers might spend more time attending to individual subevents compared to English speakers or might shift attention between subevents if they are undecided or uncertain about how to package them (see Griffin & Spieler 2006; Russo & Rosen, 1975 on such effects on eye gaze).

We also probe what effects, if any, the way that young children in both language groups talk about causative events has on the way that they examine these complex events. We investigate these processes in 3-year-old children because, as mentioned previously, this is the earliest age at which cross-linguistic differences in event description have been documented. To capture potential developmental differences in language-specific patterns of event description and eyegaze, we also include 4-year-old children. To the extent that eyegaze patterns observed during event description reflect the task of language production, we expect to see task-specific effects on attention, even in young learners. A comparison between free-viewing and language production tasks has the potential to inform the current debate about the relation between language and event cognition: to the extent that the two tasks yield distinct patterns of event inspection, this would offer evidence against the possibility that one's native language shapes the way events are parsed and processed (see Papafragou et al., 2008, for a similar argument).

Moreover, as with adults, of interest is whether child speakers of different languages who

provide the same information about causatives appear to gather that information in similar ways

as they prepare to speak. If, as in the case of motion, language-specific ways of encoding

causatives are already in place in children as young as 3 or 4 years, a particularly interesting

possibility is that (as predicted for adults) children's eye movements may reveal sensitivity to

event codability differences between the languages even if one holds utterance content constant.

Such effects, however, are expected to be relatively weak, since children are presumably less

familiar with the typologically preferred event packaging options in their language compared to

adults.

Methods

Participants

The final data sample included native speakers of English and Greek in three age groups:

3-year-olds (English: $n=20$, mean age 3;5, range 3;0–3;11; Greek: $n=20$, mean age 3;7, range

3;2–3;11), 4-year-olds (English: $n=20$, mean age 4;6, range 4;1–4;11; Greek: $n=20$, mean age

4;7, range 4;0–5;0), and adults (English: $n=20$; Greek: $n=19$). English-speaking children were

recruited through preschools in Newark, DE ($n=38$) and Philadelphia, PA ($n=2$). Greek-speaking

children were recruited through public ($n=9$) and private ($n=31$) preschools in and around

Ioannina, Greece. Children had no parent-reported history of visual, cognitive, or language

impairments. English-speaking adults were students at the University of Pennsylvania. Greek-

speaking adults were students at the University of Athens ($n=17$) or the University of Delaware

($n=2$). All Greek speakers were given instructions and tested in Greek by a native speaker of the

language. The two Greek adults from the University of Delaware were tested within 2 weeks of

returning from a 3-month trip to Greece, by a native speaker of Greek that they knew and with whom they were accustomed to conversing primarily in Greek. Adults received course credit or a small monetary compensation for participation. Data from an additional eight preschoolers and five adults were excluded from the analysis for the following reasons: unwillingness to cooperate ($n=1$ preschooler), experimenter error ($n=4$ preschoolers), or significant trackloss during stimulus viewing ($n=3$ preschoolers, $n=5$ adults; see "Analysis of eye movement data" for trackloss criteria).

Materials

The stimuli consisted of still clip-art images that depicted the midpoint of various events. Twelve target items were created that depicted agent-driven causative events that resulted in a change of location for the theme object. Each target image captures the moment just after an animate agent has used some instrument (a tool or body part) to interact with an inanimate theme object (the Means subevent) in a way that has caused the theme to move toward a visible goal (the Result subevent). A sample target item is provided in Fig. 1: in this event, the boy has used his fist to punch the soccer ball (Means), thereby sending the soccer ball (in)to the basket (Result). Event elements were assigned to stimuli in combinations that were meant to be unexpected, e.g., a boy punching a soccer ball instead of kicking it; a boy hitting a basketball (rather than a volleyball) over a volleyball net. The goal of this aspect of the stimulus design was to increase the felicity of mentioning each depicted item and, as a result, to elicit more specific event descriptions from participants (e.g., "The boy hit the basketball over the net" rather than "The boy was playing volleyball"). This goal to balance the salience of relevant aspects of the event scenes also motivated our choice to present stimuli as static scenes: omitting motion from

the events allowed us to reduce eye movements that might be driven by low-level changes to

perceptual salience (e.g., movement of objects capturing attention) rather than compilation of

conceptual event representations. Previous work (often involving on-line measures) has

demonstrated that both children (Göksun, et al., 2011; Nappa, Wessell, McEldoon, Gleitman &

Trueswell, 2009; Papafragou et al., 2006; Wilson, Unal, Trueswell & Papafragou, in prep) and

adults (Bock, Irwin & Davison, 2004; Griffin & Bock, 2000; Hafri et al., 2013; Wilson et al.,

2011) compile sophisticated intuitions about events on the basis of static scenes in tasks

involving nonlinguistic event representation as well as picture description. Moreover, it has been

demonstrated that in adults static depictions of implied motion can activate the same cortical

regions involved in the visual processing of motion that dynamic events do (Kourtzi &

Kanwisher, 2000).

(Figure 1 about here)

Agents, instruments, themes, and goals for each target event were represented by distinct

visual elements, and clipart images were constructed such that each of these event elements was

separated spatially from the others, allowing looks to each item to be distinguished in the

analysis of eyegaze data.

Three target stimuli were ultimately excluded from the final data analysis: One was

inadvertently left out of the lists of stimuli presented to adult participants, and so no data were

collected for it. The other two elicited event descriptions that suggested that study participants

had not interpreted the image in the way that experimenters intended.

The twelve filler items were clip-art images that depicted animate agents involved in events in which the agent did not cause a theme object to move from one place to another (e.g., a snail and a rhinoceros playing trumpets).

When presented on the Tobii screen, stimulus images were 31.4 x 25.1 cm (29.3 x 23.6 deg) for adults and 23.6 x 16.7 cm (22.2 x 15.9 deg) for preschoolers.[3]

Procedure and Experimental Design

Participants were randomly assigned to one of two tasks: half of the participants in each language and age group were assigned to a Linguistic task; the other half to a Nonlinguistic task (n=9 Greek adults, n=10 from each of the other groups). Demographic information for children in each condition were as follows: Linguistic task: n=10 English-speaking 3-year-old children (6 female, 4 male; mean age 3;4), n=10 English-speaking 4-year-old children (3 female, 7 male; mean age 4;7), n=10 Greek-speaking 3-year-old children (4 female, 6 male; mean age 3;8), n=9 Greek-speaking 4-year-old children (2 female, 8 male; mean age 4;9). Nonlinguistic task: n=10 English-speaking 3-year-old children (5 female, 5 male; mean age 3;6), n=10 English-speaking 4-year-old children (4 female, 6 male; mean age 4;6), n=10 Greek-speaking 4-year-old children (6 female,4 male; mean age 3;6), n=10 Greek-speaking 5-year-old children (5 female,5 male; mean age 4;5).

Each participant was run individually at his/her university campus or preschool. Stimulus presentation and data collection were carried out using a Tobii 1750 remote eyetracking system, which collected binocular eye movement data at a consistent 50 Hz sampling rate. Participants were seated approximately 60 cm from the Tobii screen. To limit head turns and whole-body movement during the session, preschoolers sat unconstrained in a car seat firmly attached to a

chair. All participants were calibrated using the ClearView software's default 5-point calibration scheme.

After completing the calibration routine, participants were given instructions specific to their task. Instructions were presented in each participant's native language. Children and adults were given the same instructions, with minor age-appropriate changes in wording. In both of the tasks, participants were informed before stimulus presentation began that they would be viewing a set of pictures showing "people and animals doing things." All participants were asked to pay careful attention to the pictures in preparation for an upcoming memory task/game.[4] Participants in the Linguistic task were additionally asked to "describe the picture" when they heard the beep that accompanied it (For preschoolers: "Tell me what's happening in the picture"). Overt linguistic encoding of events by participants in the Nonlinguistic task was discouraged: participants in this task who began to give descriptions were asked to "please watch quietly."

Stimulus presentation in both tasks followed the same progression: during an initial encoding phase, participants viewed a sequence of still images presented in a fixed semi-random order. Adult participants were presented with 11 target items (two of which were excluded from the data analysis) and 12 fillers; children were presented with a subset of these items, 8 target pictures (one of which was excluded from the data analysis) and 8 fillers. In both tasks, a beep sounded with the onset of each stimulus image. Presentation of stimulus items was experimenter-controlled for adults, with a target display time of 5 sec per picture during the encoding phase. Stimulus presentation during the encoding phase was automated for preschoolers, and each picture was displayed for 9 sec. Pilot testing of the stimuli revealed that children took longer than adults to begin their descriptions of stimulus items. The extra time allowed for stimulus presentation for children permitted them to complete their event descriptions before the pictures

disappeared from the screen. For both age groups, a fixation cross appeared before each stimulus item. The location of the cross was fixed individually for each stimulus item so that it would appear approximately equidistant from, but not in the same location as, the agent, instrument, theme, and goal of the following picture. Participant responses were provided verbally, and sessions in which participants completed the Linguistic task were audio recorded to facilitate accurate collection of event descriptions.

Data Coding and Analysis

Coding of event descriptions. Event descriptions collected from participants in the Linguistic task were transcribed and hand-coded by native speakers of the language under consideration. Descriptions of target items were assessed for mention of the two subevents of each causative target. Words and phrases that specifically encoded either the agent's activity or the nature of the agent's contact with the theme object were coded as mention of the Means subevent. Words and phrases that identified either what happened to the theme object as a result of that activity or to some underspecified goal-oriented intention on the part of the agent (e.g., "put") were coded as mention of the Result subevent. Using these codes, each utterance was then classified as including specific information about only the Means subevent of a depicted causative event (Means Only), only the Result subevent (Result Only), both the Means and Result subevents (Means+Result), or neither subevent of the depicted causative (Neither). For example, for the target event depicted in Fig. 1, participants used verbs like "punch"/"hit" (Greek "htipo") and "throw" ("rihno") to describe the Means subevent, and verbs like "put" ("vazo") and prepositional phrases like "into the basket" ("sto kalathi") to describe the Result subevent. An utterance like sentence (6a) would be classified as Means Only, and sentence (6b) would be

classified as Result Only. An utterance like sentence (6c) would be classified as an event

description in which both the Means and the Result were mentioned. Finally, an utterance like

(6d) would be classified as an event description that included information about neither of the

causative subevents.

(6)	a. The boy punched a soccer ball.	Means Only
	b. The boy put the soccer ball into the basket.	Result Only
	c. The boy punched a soccer ball into a basket.	Means+Result
	d. The boy is playing ball.	Neither

Analysis of eye movement data. Eye movement data from target items were analyzed to

assess the effects of task, age group, and language background on inspection of causative

subevents. Data samples (50 per second) were time-locked to the onset of each stimulus item,

and analyses were performed on raw eyegaze coordinates from each sample. Trackloss was

determined separately for each eye by Tobii's ClearView software; our data set includes samples

for which the system is certain it has recorded correct coordinates for at least one eye. Gaze

coordinates were taken from eyes with no trackloss (or from an average of both eyes, if neither

had trackloss). Target trials with global trackloss of >30% were excluded from the analysis

($n=27$ for 3-year-olds, $n=23$ for 4-year-olds, $n=8$ for adults). Trials were excluded evenly across

target items for each age group. Eight participants who had more than four excluded target trials

were excluded from the analysis; as described previously, these participants were replaced in the

study design.

Two spatial scoring regions were defined for each target image that corresponded to the Means and Result subevents of the depicted causative. These regions were defined on the basis of independent norming by adult raters from each language group (*n*=10 English speakers, *n*=10 Greek speakers) who were asked to draw rectangles around the part of each target picture that showed "the Result of the event" and the part that showed "the Central Action or Means." Spatial scoring regions used for Means and Result subevents that were used in the eyegaze analysis were based on the most common region assignments across items: roughly, looks to the instrument used by the agent to interact with the theme object were interpreted as looks to the Means subevent, and looks to the goal object were interpreted as looks to the Result subevent. For the event depicted in Fig. 1, looks to the boy's fist were interpreted as looks to the Means subevent, and looks to the basket were interpreted as looks to the Result subevent. Although theme objects (the ball, in Fig. 1) are conceptually critical to both causative subevents, our raters were more likely to omit them from each region than to include them. Accordingly, looks to themes are not reported here: themes were not included in either Means or Result regions, and were excluded from these regions in cases of overlap. The head and upper torso of the agent of each event were also excluded from Means and Result regions and comprised an independent region of interest; looks to this Agent region are not discussed here because they do not bear on the question under discussion.[5] Means and Result regions never overlapped in spatial coordinates. Exact sizes of Means and Result regions for each target image can be found in the Appendix. On average, Means regions subtended 12.9 x 11.6 deg visual angle for adults and 9.7 x 7.8 deg for preschoolers, and Result regions subtended 12.0 x 10.7 deg visual angle for adults and 9.0 x 7.2 deg for preschoolers. The increase in the distance between scene components for adults (see footnote 3) was largely offset by the necessary increase in the size of the AOIs drawn

around those scene components. Eyegaze data are reported as the proportion of samples

(averaged across subjects) for looks within these predefined regions of interest (out of all

looking), averaged into blocks of 1 sec. Any looks within a region were included in the analysis,

regardless of duration.

Statistical analyses. The reliability of trends observed in the data was tested using

multilevel mixed logit modeling with crossed random variables for Subjects and Items (after

Baayen, Davidson & Bates, 2008; Barr, Levy, Scheepers & Tily, 2013). Fixed factors that varied

within subjects and interactions between those factors were included as random slopes in Subject

and Item effects structures when they were not perfectly correlated with the intercept. All *p*

values reported for factors within analyses are vs. an empty model with no fixed effects.

Results and Discussion

Causative Event Descriptions

Event descriptions that included no information about either of the causative subevents

depicted in the corresponding target event ("Neither" responses, 38% of all event descriptions

across age and language groups[6]) were omitted from further analyses. In most of these cases,

participants described the causative event at a global level rather than identifying individual

subevents, for example, describing the event in Fig. 1 as "a man playing ball."

Table 1 provides information about the causative subevents that participants *did* mention

in their event descriptions. There were both similarities and differences in patterns of event

description across language and age groups: within both language groups, adults were more

likely than children in both age groups to mention both subevents as opposed to just one.

However, English speakers (both children and adults) were more likely than Greek speakers to

mention both subevents. These trends were confirmed by multi-level modeling of categorical

values at the trial-level for number of subevents mentioned in event descriptions (1, 2) with Age

(Adult, 4-year-old, 3-year-old) and Language (English, Greek) as first-level fixed factors. The

best-fitting model (p<0.01; Table 2) includes main effects of both Age and Language, but no

interaction between the two. The effect of Age was driven by differences between adults and

children in each language group; there were no differences between 3-year-olds and 4-year-olds.

(Table 1 about here)

(Table 2 about here)

When participants did provide information about both subevents, they mentioned the

Means first in 121 of 123 descriptions. The two Result-first descriptions were both produced by

Greek-speaking adults. English speakers tended to encode the Means subevent in a verb and the

Result in a post-verbal prepositional phrase (79/80 descriptions that mentioned both subevents),

as in the sentence in (7a). When the verb that Greek speakers used in their Means-first event

descriptions permitted them to include a resultative phrase, as in the case of the verb "rihno"

(English "throw") (27/41 descriptions), they also produced sentences like (7a). When it did not

(14/41 descriptions), Greek speakers provided two-clause descriptions, the first including a verb

that encoded the Means subevent and the second a verb that encoded the Result subevent, as in

(7b).

(7)   a. A person just punched a soccer ball into a basket.   Adult

         b. Enas antras htipai mia bala gia na ti vali sto kalathi.   Adult

'a man is-hitting a ball so as to put-it into the basket'

In the rest of the relevant event descriptions, participants mentioned only one causative subevent. As seen in Table 1, adult speakers of both languages were more likely to mention just the Result subevent (26/39 event descriptions), whereas children from both language groups were more likely to mention just the Means (65/102 event descriptions). In fact, mentioning just the Means subevent was the preferred response for preschoolers as a group. The sentences in (8), which encode just the Means subevent, and (9), which encode just the Result subevent, were provided as descriptions of the target event in Fig. 1.

(8)   a. He's punching a soccer ball.   4-year-old

        b. Anthropos pu erikse ti bala.   4-year-old

          'man who threw the ball'

(9)   a. He's trying to get the soccer ball into the basket.   3-year-old

         b. Prospathi na vali afti ti bala mesa edo.   4-year-old

          'tries to put this ball in here'

Table 3 provides details for the best-fitting model of categorical values at the trial-level for single subevent mentioned (Means, Result) with Language (English, Greek) and Age (Adult, 4-year-old, 3-year-old) as (between-subjects) first-level fixed factors. The model includes an effect of Age for 3-year-olds vs. adults ($p<0.05$), but no effect of Language and no interaction between the two variables. There was no significant difference between 3-year-olds and 4-year-olds as both were producing a high number of event descriptions that mentioned only the Means.

(Table 3 about here)

In sum, analyses of participants' utterances point to both similarities and differences in the density of information that speakers from different language and age groups provided about causative events. As expected, adult speakers of both English and Greek were more likely to mention both causative subevents rather than just one, thereby providing detail about both the Means and the Result of these complex events. Conversely, preschool-aged speakers of both languages were very likely overall (and more likely than adults) to mention only one subevent in their descriptions of causative events. These developmental differences are reminiscent of previous research that has found non-adult levels of informativeness in children's event descriptions (e.g., Bunger et al., 2012; Ji et al., 2011; Özyürek et al., 2008; Papafragou & Selimis, 2010).

Despite these similarities across language groups, as we anticipated, English speakers were more likely than Greek speakers to mention both causative subevents when describing our target events. Moreover, this difference was already evident in preschool-aged speakers of the two languages. Thus by age 3, children are already sensitive to some language-specific patterns of event description, even for complex events like causatives (see Introduction for similar findings with motion events). Furthermore, in situations in which both subevents were mentioned, their packaging was more consistent in English compared to Greek: English speakers virtually always used a Means verb (e.g., *punched*) accompanied by a Result prepositional phrase (*into the basket*), while Greek speakers used such structures about 60% of the time and split Means and Result encoding over two separate clauses (and verbs) about 40% of the time. Thus

both in terms of both quantity and consistency of causative encoding, English appears to follow a single typologically dominant pattern whereas Greek shows a more mixed pattern.

Finally, we found that, when speakers mentioned only one causative subevent, preschoolers were more likely to mention the Means whereas adults were more likely to mention the Result. There are several possible reasons for this difference. One possibility is that children tended to mention only Means subevents because of constraints on linguistic encoding: i.e., they may have planned event descriptions that included both subevents, as adults tended to do, but constraints on linguistic processing (e.g., the length or syntactic complexity of the required utterance) prevented them from producing both subevents. According to this view, preschoolers were led to drop the subevent that adults tend to mention *second* when they produce full (2-subevent) causative descriptions. Alternatively, children and adults might differ in terms of conceptual event encoding: adults, but not children, might have thought that the Results depicted in our pictures were central to the core of a causative event. Consistent with this hypothesis, prior studies have shown that adults have a strong bias for Results when categorizing or learning about causative events (e.g., Papafragou & Selimis, 2010) but children present a mixed picture, taking the activity of an agent (Casasola & Cohen, 2000), the result (Bunger, Baier & Lidz, 2009), or neither/both (Papafragou & Selimis, 2010) to be the core of a causative event. Because there were several differences among these past studies in terms of the stimuli and methods used, this issue remains ripe for further investigation.

Detailed Analysis of Eye Movements: Adult Data

Given the developmental and cross-linguistic similarities and differences in the linguistic content that speakers provide when they describe causative events, we next want to look at eye

gaze data within the Linguistic task to ask whether preparing this content also affects the way

speakers gather information from the visual world while they describe causatives. We are

specifically interested in testing two possibilities. According to the first possibility, thinking for

speaking would proceed identically for speakers of different languages as long as people produce

surface-identical event descriptions: for instance, regardless of native language, people should

attend to event components they plan to mention in the order in which they plan to mention them

(see, e.g., Griffin & Bock, 2000; cf. Slobin, 1996). Beyond this broad similarity, speakers of

different languages who describe events in the same way might diverge somewhat in their

patterns of event inspection depending on how closely their linguistic choices correspond to the

preferred typological pattern in their language (an idea consistent with Slobin, 2004). To test

these possibilities, we follow prior (single-language) production studies that assess subsets of the

data, making splits on the basis of shared form or content of utterances (Bock, Irwin, Davidson

& Levelt, 2003; Gleitman et al., 2007; Griffin & Bock, 2000; Griffin & Oppenheimer, 2006).

Beginning with the adult data, we analyze eyegaze only for trials on which both causative

subevents were mentioned. Overall, these event descriptions corresponded to the majority of

adult responses but represented a more consistent, typologically dominant pattern in English

compared to Greek (recall that English speakers were more likely than Greek speakers to

produce such strings and to package them within a single clause). As a further step, to understand

whether the eyegaze patterns in the Linguistic task are linked to the process of language

production or whether they are representative of baseline/nonlinguistic patterns of interest in the

causative subevents, we also compare adult eyegaze across the Linguistic and Nonlinguistic

tasks.

Adult eyegaze: Linguistic task. Figure 2 shows eyegaze patterns for adults in the

Linguistic task who mentioned both causative subevents in their event descriptions. Data in Fig.

2 have been aligned by speech onset on each trial: data are shown for the time period that

includes the second before speech onset (-1 to 0 s) and the first three seconds after speech onset

(0 to 3 s). Two Greek adults from the Linguistic task for whom speech onsets could not be

collected have been excluded from the data set reported in this section. Three additional trials

have been excluded because speech onset occurred less than 3 s before the target image

disappeared from the screen (*n=2* trials from English-speaking adults, *n=1* trial from Greek-

speaking adults).

Informal inspection of Fig. 2 suggests that adults from different language backgrounds

inspected target causative events in similar ways while producing event descriptions that

included information about both Means and Result subevents. Adults from both language groups

showed more attention to Means regions just before the onset of their event descriptions and for

about a second after speech onset, and directed an increasing amount of attention to Result

regions during the last 1–2 s of stimulus viewing shown here. This order of preference for the

two regions directly corresponds to the order in which speakers mentioned the subevents in their

event descriptions: Means first, and then Result. Multilevel mixed modeling confirmed that there

were no reliable effects of Language on looks to Means or Result regions for any of the four

analysis windows. Elogit-transformed proportions of looks to Means and Result regions were

modeled separately within four 1-s windows beginning 1 s before speech onset and ending 3 s

after speech onset, with Language (English, Greek) entered as a first-level fixed factor. For every

window, the model including Language as a fixed factor did not provide a better fit for the data

than an empty model with no fixed factors.

(Figure 2 about here)

Summarizing, in the Linguistic task, the way that adult viewers directed their attention to causative events as they described them revealed speech planning effects on attention regardless of language: speakers encoding both Means and Result subevents attended to the corresponding event components in the order in which they planned to mention them. These data are consistent with prior work on English showing order-of-mention effects for individual event participants (e.g., Agents/subjects and Patients/objects) on attention during language production (Griffin & Bock, 2000; also Bock et al., 2003; Gleitman et al., 2007). Our data expand this earlier work by showing that such order effects emerge cross-linguistically, as long as speakers employ comparable surface arrangements of linguistic information.[7] Beyond these broad similarities, our data offer no support to the idea that identical linguistic outputs might result from different speech planning routes—themselves detectable in attention patterns—depending on the codability of events in one's native language (an idea indirectly suggested in Slobin, 2004). Despite the fact that the Means+Result packaging option is not as frequent or consistent in Greek as it is in English, production of this type of event description in both languages is accompanied by identical shifts of attention to Means and Result subevents.

Adult Eyegaze: Linguistic vs. Nonlinguistic tasks. To confirm that the adults' eyegaze patterns in the Linguistic task are specific to speech planning processes, we would need to determine that these patterns are in some respect distinct from those observed when participants did not have to prepare to speak. With this goal, in this section, we compare adult eyegaze patterns across the Linguistic and Nonlinguistic tasks.

Figure 3 shows the average proportion of eyegaze that adults in the two tasks allocated to our Means and Result regions of interest, now aligned by onset of the target picture rather than onset of speech. For brevity, only the first 4 s of stimulus viewing are presented. We tested the effects of Language and Task on adult eyegaze patterns using multilevel mixed modeling as described previously. Elogit-transformed proportions of looks to Means and Result regions were modeled separately within four 1-s windows beginning from the onset of target images, with Language (English, Greek) and Task (Linguistic, Nonlinguistic) entered as first-level predictors. Modeling revealed main effects of both Language and Task (Fig. 3) on the way that adults examined target stimuli.

(Figure 3 about here)

When eyegaze data are collapsed across Tasks, English-speaking adults directed their attention to Means regions earlier than Greek-speaking adults (Fig. 3A). Language was a reliable predictor of looks to Means regions in the first analysis window: during the first second after the onset of the stimulus image, English-speaking adults directed more attention to Means regions than Greek-speaking adults did (Table 4, $p<0.05$). There were no significant effects of Language on looks to Means regions in the remaining three analysis windows, and no effects of or interactions with Task in any analysis window.

Moreover, in the Linguistic task, both groups of adults show an early delay and a late increase in their attention to Result regions compared to patterns of eyegaze by adults in the Nonlinguistic task (Fig. 3B). Task was a reliable predictor of looks to Result regions by adults for the analysis windows that included eyegaze data from 1–2 s and 3–4 s after the onset of the

stimulus image (Table 4). In the 1–2-s analysis window, adults from both language groups in the

Nonlinguistic task directed more attention to Result regions than did those in the Linguistic task

(p<0.01). In the 3–4-s analysis window, this pattern was reversed: adults from both language

groups in the Linguistic task directed more attention to Result regions than did those in the

Nonlinguistic task (p<0.05). There were no significant effects of Task on looks to Result regions

in the other two analysis windows, and no effect of or interactions with Language in any analysis

window.

(Table 4 about here)

Our comparison between eye movements in the Linguistic task and the Nonlinguistic task

reveals that the process of language production shifted adult speakers' attention away from their

baseline preferences. Specifically, in the Linguistic task, adults from both language groups who

were planning event descriptions that included information about both the Means and the Result

of a causative event turned their attention to Result regions later than adults in the Nonlinguistic

task did. This delay in attention to Result regions is presumably linked to the fact that speakers

tended to mention those subevents second (i.e., after Means subevents) in their event

descriptions. Thus these data support the conclusion that the attention patterns observed in the

Linguistic task are not simply attributable to overall preferences in inspecting causative events

but that the process of language production directs cognitive resources in ways that are relevant

to the preparation of to-be-mentioned event features (cf. Griffin & Bock, 2000; Papafragou et al.,

2008). The presence of distinct patterns of attention allocation in the two tasks is consistent with

the view that aspects of event cognition are independent from one's native language (see

Papafragou et al., 2008).

Our data did uncover an effect of language background on looks to Means regions across

tasks. One might attribute this effect to cross-linguistic differences in the conceptual encoding of

causative events since English-speaking adults were numerically more likely to mention Means

subevents than were Greek-speaking adults in their overt descriptions. Note, however, that

although they directed less attention to Means regions than English speakers did during the first

second of the viewing period, Greek speakers were not directing more attention to Result regions

(Fig. 3B). It thus remains an open possibility that the observed asymmetry was due to the fact

that speakers were covertly encoding the events linguistically (at least to some extent) during the

Nonlinguistic task. We revisit the significance of this result in the General Discussion below.

Detailed Analysis of Eye Movements: Preschool Data

Next, we turn to an analysis of eyegaze by preschoolers, asking 1) whether, like adults,

they show a link between attention and production for causative events during the Linguistic

task, and 2) whether this link remains stable for utterances of comparable content regardless of

native language, as we found for adults. We split the eyegaze data to reflect two broad types of

event descriptions, grouping trials into (i) those on which preschoolers failed to mention the

Result of the causative event (Means Only in example (6)), and (ii) those on which preschoolers

did mention the Result (Result Only and Means+Result in example (6)). This way of splitting the

data gives an approximately even partition, since Means Only responses are the most prevalent

description type in children across both language groups and the frequency of this type of

response is equivalent to the frequency of the other two description types combined (Table 1).

Finally, as for adults, we ask whether preschoolers' attention patterns in the Linguistic task

reflect the task of language production rather than preferences that also appear in the

Nonlinguistic task. Eyegaze from preschoolers in each task is collapsed across age groups

because we found no age-related differences in children's linguistic output.

Preschooler eyegaze: Linguistic task. Beginning with the Linguistic task, Figure 4

shows eyegaze patterns for preschoolers who mentioned either just the Means subevent or the

Result subevent (perhaps in combination with the Means) in their descriptions of causative

events. As in Fig. 2, the data in Fig. 4 have been aligned by speech onset on each trial: data are

shown for the time period that includes the second before speech onset (-1 to 0 s) and the first

three seconds after speech onset (0 to 3 s). One Greek preschooler from the Linguistic task for

whom speech onsets could not be collected has been excluded from the data set reported in this

section, and one additional trial from an English-speaking 3-year-old has been excluded because

speech onset occurred less than 3 s before the target image disappeared from the screen.

(Figure 4 about here)

Informal inspection of Fig. 4 suggests that Greek- and English-speaking preschoolers

showed similarities in the way that they inspected causative events. Before they began to speak,

both groups of preschoolers appear to have directed more attention to Means regions than to

Result regions, regardless of the subevents they ended up mentioning. After this point, however,

preschoolers began to direct their attention differently depending on the type of event description

they produced: specifically, preschoolers directed more attention to Means regions on those trials

in which they mentioned just the Means of target causatives (Fig. 4A, 0–2 s), and more attention

to Result regions for those trials on which they mentioned Results (Fig. 4B, 2–3 s). We tested the reliability of these effects using multilevel mixed modeling as described previously. Elogit-transformed proportions of looks to Means and Result regions were modeled separately within four 1-s windows beginning 1 s before speech onset and ending 3 s after speech onset, with Language (English, Greek) and Production Type (Means Only, Result) entered as first-level fixed factors. As anticipated, the analysis showed that Production Type was a reliable predictor of looks to Means regions in the second and third analysis windows (Table 5, both p<0.05), and looks to Result regions in the final analysis window (Table 5, p<0.01).

Despite these similarities, there was a significant difference in the way preschoolers from different language groups inspected causative events. Although both groups of preschoolers increased their attention to Result regions slightly after they began to speak, Greek-speaking preschoolers were more likely than English-speaking preschoolers to attend to Means regions regardless of the content of their utterances (Fig. 4A, 1–2 s). Indeed, Language was a reliable predictor of looks to Means regions in the third analysis window (Table 5, $p<0.01$). No other effects or interactions were significant.

(Table 5 about here)

Summarizing, in the Linguistic task preschoolers, like adults, allocated more attention to causative event components that they planned to talk about: specifically, they directed more attention to Means regions when planning descriptions that mentioned only the Means subevent and more attention to Result regions when planning event descriptions that included Result information. This is the earliest demonstration to date of a link between patterns of attention and

speech production in children as young as 3 and 4 years of age. Furthermore, this strong link

between production and attention held regardless of children's native language.

Our data also point to a late difference between the two language groups, with Greek-

speaking preschoolers directing more attention to Means regions compared to English-speaking

preschoolers regardless of production type. This effect might be a specific outcome of speech

planning—that is, unlike adults, preschoolers may be showing language-related differences in

event inspection while planning utterances of equivalent content. Alternatively, it could be due to

broader, nonlinguistic biases in event encoding. To understand the nature of this effect, we turn

to children's eye gaze patterns during a task that did not involve producing language.

Preschooler eyegaze: Linguistic vs. Nonlinguistic tasks. To test whether eyegaze

patterns in the Linguistic task truly characterize effects of language processing or are equivalent

to nonlinguistic preferences for processing causative subevents, we next compare eyegaze from

preschoolers across the Linguistic and Nonlinguistic tasks.

Figure 5 shows the average proportion of eyegaze that preschoolers in the two tasks

allocated to our Means and Result regions of interest, now aligned by onset of the target picture

rather than the onset of speech. For brevity, only the first 4 s of stimulus viewing are presented.

An informal inspection of Fig. 5 suggests that preschoolers inspected Means and Result

subevents over time in roughly equivalent ways across both tasks and language groups.

(Figure 5 about here)

We tested the reliability of these comparisons between eyegaze by preschoolers in the

Linguistic and Nonlinguistic tasks using multilevel mixed modeling as described previously.

Eyegaze from the linguistic task remained split by production type: separate comparisons to

Nonlinguistic eyegaze were performed for Means Only event descriptions and event descriptions

that contained Result information. For each comparison, elogit-transformed proportions of looks to Means and Result regions were modeled separately within four 1-s windows beginning from the onset of target images, with Language (English, Greek) and Task (Linguistic, Nonlinguistic) entered as first-level predictors. For the comparison between Nonlinguistic trials and Linguistic trials on which participants mentioned Result information, Task was a reliable predictor of eyegaze for looks to Result regions in the third analysis window (Table 6, p<0.05): preschoolers in the Linguistic task directed more attention to Result regions 2–3 s after speech onset for those trials on which they mentioned information about the Result subevent compared to the attention they directed to Result regions during the same time period in the Nonlinguistic task. No other reliable effects of Language or Task or interactions between the two were found.

(Table 6 about here)

Our comparison between eye movements by preschoolers in the Linguistic and Nonlinguistic tasks reveals subtle effects of the process of language production on the way that preschoolers extract information from causative events. Specifically, for those trials on which preschoolers in the Linguistic task mentioned the Result subevent, they directed more attention to Result regions in comparison to preschoolers in the Nonlinguistic task. We did not find an equivalent effect of language production on looks to Means regions, perhaps because across tasks, speakers were already directing more attention to Means regions than to Results. Thus language preparation produces strong pulls on both adults' and young children's attention (cf. Griffin & Bock, 2000; Bock et al., 2003, among others).

Preschoolers' eyegaze patterns in the Nonlinguistic task did not differ across language groups, suggesting that Greek- and English-speaking children had equivalent baseline preferences for the components of our target events. This is important for two reasons. First, this finding shows that the language-specific difference we found in preschoolers' attention to Means regions in the Linguistic task cannot be attributed to extralinguistic factors, but is linked to language-specific processes of speech planning (potentially related to differences in the encoding of causatives between English and Greek; cf. Slobin, 2004). Second, the Nonlinguistic task suggests that children's conceptual representations of caused motion events are similar across different language-learning populations – a finding that bears directly on current debates about language and thought (Gentner & Goldin-Meadow, 2003; Gleitman & Papafragou, in press; Wolff & Holmes, 2011). We return to the significance of these results in the discussion below.

General Discussion

In the present study, we used a novel combination of cross-linguistic and online methods to investigate the inspection and description of causative events by Greek- and English-speaking preschoolers and adults. We aimed to highlight cross-linguistic similarities and differences in how subcomponents of causative events are encoded in language, contributing to our understanding of what might be shared vs. language-specific about the way that events are encoded in the speech of both mature and novice language users. In doing so, we wanted to broaden the empirical basis of theories of event representation beyond prior work on motion events. Additionally, we used eyetracking to probe the process that speakers go through when

selecting causative information to talk about, shedding light on the ways language production (or thinking for speaking) operates across speakers of different languages, changes with development, and interfaces with nonlinguistic cognition.

Causative Events in Language

Our findings demonstrate that the cross-linguistic encoding of causatives is sensitive to the internal structure of the events, i.e., it makes reference to the Means and Result subevents within the causal chain (cf. Jackendoff, 1990; Levin, 1993; Levin & Rappaport Hovav, 1995; Pustejovsky, 1991; Talmy, 1985). Nevertheless, there are language-specific patterns in the way that causal information is encoded: across age groups, Greek speakers in this study were less likely to include information about both causative subevents compared to English speakers. We have hypothesized that this difference is related to the fact that Greek, unlike English, places restrictions on whether both causative subevents can be encoded within a single clause, and therefore lacks a single productive event packaging (such as the option "Means V + Result PP" in English) but instead chooses from a set of possible information packaging alternatives. The fact that this cross-linguistic difference emerged across age groups suggests that by 3 years of age children are already beginning to exhibit some language-specific structural patterns in their description of complex causative events. This result adds to several studies that have shown early convergence on language-specific patterns for simpler, spontaneous motion events (Allen et al., 2007; Naigles et al., 1998; Özyürek et al., 2008; Papafragou et al., 2006; Papafragou & Selimis, 2010; Slobin, 1996; Zheng & Goldin-Meadow, 2002).

Our findings further show that, across language groups, the tendency to package information about both causative subevents increased with development, such that adult speakers

of both languages tend to mention more causative subevents than children do. Previous work has

shown that young children tend to provide less information about motion events than adults do

(e.g., Bunger et al., 2012; Özyürek et al., 2008; Papafragou & Selimis, 2010). Here, we extend

this finding to a class of more complex events. Contrary to what has been reported for motion

events, however, we found that the causative subevent that speakers were most likely to mention

differed across age groups but not across language groups. Adult speakers of both languages

were more likely to provide information about the Result of causative events than information

about the Means whereas the opposite was true of preschool-aged speakers from both language

groups (even though the precise explanation of this difference remains open; see Causative Event

Descriptions section).

Perhaps somewhat surprisingly, our linguistic data revealed no difference in the type of

event descriptions offered by 3- and 4-year-olds in either language. It is possible that a more

complete encoding of causatives needs to await developments in children's syntactic knowledge

and clause-building capacities that take place later in development. Alternatively, or additionally,

the need to make inferences about causality on the basis of the still images we presented as

stimuli in this study may have posed a challenge that limited the processing resources that were

available to devote to language production within the timeframe of each individual trial. In this

case, children's incomplete event descriptions might be due at least in part to pragmatic

limitations that result in underinformative messages and that may take longer to overcome (see

Bunger et al., 2012 for discussion).

Attention to Causative Events During Language Production

Our analysis of eyegaze patterns allowed us to explore the way that both adults and children collect information from the visual world as they prepare to organize a dynamic event into a linguistic string. One possibility, strongly suggested by prior work on English-speaking adults (e.g., Bock et al., 2003; Gleitman et al., 2007; Griffin & Bock, 2000; Griffin & Spieler, 2006; cf. also Levelt, 1989; Slobin, 1996) was that semantic content and its surface organization might guide attention allocation as one prepares to speak. Previous work has shown that when adults prioritize information differently in descriptions of motion events, they exhibit different patterns of attention allocation while planning to speak (e.g., Papafragou et al., 2008). Here, we investigated cross-linguistic attention patterns during the planning of utterances that share similar semantic content and organization. If these factors do play a critical role in guiding attention during language planning, then adults producing equivalent causative descriptions should inspect causative events similarly when planning their utterances regardless of their native language. Furthermore, assuming some basic degree of continuity in the architecture of the developing production system, such language-driven effects on attention should also surface in children speaking different languages (despite other issues in attention control that children may face; e.g., Scerif et al., 2005). A different, potentially complementary, possibility (inspired by Slobin, 2004; cf. MacDonald, 2013) was that cross-linguistic differences might affect speech planning even when speakers end up describing events similarly, as long as the events are more codable in one language compared to another. In all cases, it was of further interest to test whether speech planning would have distinct effects on attention compared to other tasks such as free viewing that do not necessarily mobilize linguistic resources (Bunger et al., 2012; Papafragou et al., 2008).

Most of these expectations were confirmed by our results. First, adults across language groups gathered information from causative events during the process of language production in ways that reflected the semantic content of their utterances (Means+Result) and its ordering in their sentences. Children who mentioned similar kinds of information in their event descriptions exhibited similar shifts in their patterns of attention to causative subcomponents during language planning, and, critically, these shifts in attention were tied to the content of their utterances. Specifically, children directed more attention to Means regions of our causative events when they produced event descriptions that included only Means information, and more attention to Result regions when they mentioned Result subevents.

Second, the observed effects on attention were task-specific: across language groups, adults in the Linguistic task who were preparing event descriptions that included information about both causative subevents directed their attention to regions of our target events that depicted Results later than did adults in the Nonlinguistic task. This delay in attention to Results is consistent with a strategy in which speakers direct their attention to event components in the order that they plan to talk about them (cf. Griffin & Bock, 2000): in this study, when adult speakers mentioned both causative subevents, they were more likely to mention Means subevents before Results. Preschoolers in the Linguistic task also demonstrated a shift in attention from their baseline preferences for the causative subevents: when they mentioned Result subevents, they directed more attention to Result regions than did preschoolers in the Nonlinguistic task.

Taken together, these two findings strengthen the conclusion that the act of speech planning changes the way that speakers examine events, with more attention being allocated to to-be-mentioned components (e.g., Bock et al., 2003; Gleitman et al., 2007; Griffin & Bock,

2000). These findings further demonstrate for the first time that the coupling of overt attention allocation with patterns of event description holds across adults from different language groups and extends to preschool-aged speakers. Just like adults, children as young as 3 or 4 years of age in this study directed more attention to to-be-mentioned event components, and their attention during language planning differed from the attention they directed to the same event components in a nonlinguistic task.[8] This pattern of results is not consistent with an account in which the salience of various event components drives preschoolers' attention during language planning. Instead, these findings reveal that thinking for speaking processes (Slobin, 1996) guide overt attention during language production cross-linguistically in young preschoolers (cf. Bunger et al., 2012). They also suggest that the link between speech planning and attention is likely to be a strong architectural feature of the developing production system.

One hypothesis that received limited support from our data was the possibility that typologically preferred event encoding patterns shape frequent online decisions about how to package events into sentences (Slobin, 2004) and thus might lead to more efficient assembly of event information. Specifically, data from children, but not adults, in this study were consistent with this hypothesis. When adults produced descriptions that contained both Means and Result information, there was no evidence that English speakers extracted the relevant information from scenes more fluently compared to Greek speakers, despite the fact that such complex descriptions were more representative of English than of Greek typology. It might be that the differences in the codability of causatives between English and Greek were not salient enough; alternatively, or additionally, the potential planning differences may be too subtle to be detected in the analysis of attention patterns we present here.

However, eyegaze patterns in the child data did reveal an asymmetry that is consistent with this hypothesis: Greek learners in the Linguistic task directed more attention to Means regions compared to English learners regardless of the type of description they produced. Since there were no language-specific differences in the Nonlinguistic task among English- and Greek-speaking children of this age, this late difference suggests that speakers of the two languages were drawing on different strategies for packaging conceptual information in linguistic structures during the process of language production. The fact that Greek-speaking children were more likely than their English-speaking peers to attend to the Means during communication might reflect the relative difficulty of packaging both Means and Results into a compact linguistic structure (notice also that Greek speakers were overall numerically more likely to provide Result-Only descriptions of caused motion events; see Table 1). At present, this effect needs to be treated with caution, since it has no counterpart in the adult data. Nevertheless, if confirmed, it would show that children that have developed language-specific (but not yet adult-like) patterns of encoding causatives experience selective difficulties in planning event descriptions depending on the codability of these events in their language.

The Nonlinguistic Apprehension of Causative Events

The eyegaze data from the Nonlinguistic task make two further contributions that are themselves of theoretical interest. First, they add to the fast-developing literature on how humans parse and interpret dynamically unfolding events of various types (Zacks & Shipley, 2008). In particular, they suggest that event perception extracts event components that are relevant for language (here, the Means and Result of a causal chain), an assumption supported by other recent studies (Bunger, Papafragou & Trueswell, 2013; Dobel, Gumnior, Bölte & Zwitserlood,

2007; Hafri et al., 2013; Webb, Knott & MacAskill, 2010; Wilson, et al., 2011). Second, our

results throw light on the relation between language and cognition, a topic that remains hotly

contested (see Introduction). Specifically, the fact that attention allocation patterns in both adults

and children diverged across linguistic and nonlinguistic tasks supports the position that

conceptual representations can be formed independent from language (see also Papafragou et al.,

2008; Trueswell & Papafragou, 2010). For children, the nonlinguistic data offer concrete

evidence for the idea that learners' emerging event representations in language build on

underlying conceptual event representations that are shared across members of different

language communities (see Gleitman & Papafragou, in press, for similar evidence from other

domains).

One aspect of our data, however, suggests a language-specific effect on event perception.

Recall that, when adult eyegaze data were collapsed across the Linguistic and Nonlinguistic

tasks, English speakers were shown to direct their attention to Means regions earlier than Greek

speakers did, and this effect of language did not interact with task (Linguistic vs. Nonlinguistic).

Since English-speaking adults were also numerically more likely to mention Means subevents

than Greek-speaking adults were (perhaps because the grammar of English makes it easier for

English speakers to encode both causative subevents in a single sentence), one possibility is to

interpret this asymmetry as an effect of linguistic relativity on conceptual encoding: Greek

speakers are less likely to encode Means, and this extends to situations in which no (overt)

linguistic encoding is required (cf. Kersten et al., 2010 for a similar account of motion event

classification by adult speakers of English and Spanish). The asymmetry does not extend to

Results since the two language groups encode Result subevents with the same frequency (Table

1).

An alternative explanation for the asymmetry in early looks to Means across adult speakers of English and Greek is that viewers covertly generated some form of linguistic encoding of the events for the purpose of remembering them, with English speakers generating Means encodings more consistently than Greek speakers. Interestingly, such strategic uses of language to support memory have generally been assumed to be absent in children before age 6 (e.g., Dessalegn & Landau, 2013), and this is supported by our own data. Previous eye tracking studies comparing Greek- and English-speaking adults' apprehension of motion events during a memory task have also found linguistic intrusions, especially under conditions of high cognitive load (Papafragou et al., 2008). Further tests showed that these intrusions reflected online use of language to recode the visual stimuli rather than more stable asymmetries in the way speakers of different languages conceptualized events: language-specific attention patterns disappeared under verbal interference that blocked access to the linguistic code but persisted when interference was nonverbal (Trueswell & Papafragou, 2010). Similar tests would help determine the nature of the observed language-specific difference in eyegaze during the present Linguistic task.

Final Summary

This study breaks new ground by providing insight into the cross-linguistic and developmental factors that affect the language production system and the attention mechanisms that it employs. As reviewed in the Introduction, previous studies have demonstrated developmental and cross-linguistic influences on the way that speakers view and describe spontaneous motion events. Here we expand our understanding of the way that speakers evaluate and describe complex events to include events of caused motion, as well as cross-linguistic

comparisons of children as young as 3 years of age. Our findings provide evidence from both

event description and patterns of eyegaze that children as young as 3 years of age are sensitive to

the way that adult speakers of their own language tend to describe causative events. Broadly,

moreover, our findings provide evidence that, as long as they are producing equivalent linguistic

output, speakers across language and age groups show fundamental similarities in the way that

they plan and execute the process of language production.

Footnotes

[3] All adults were presented with stimuli using E-Prime software and all preschoolers were presented with stimuli using ClearView software. Modification of images for presentation in ClearView resulted in the export of stimuli with an aspect ratio that was different from the images shown to adults, but with no noticeable image distortion. No comparisons of eyegaze patterns across age groups are reported in the current set of findings.

[4] Stimuli and procedure for the Memory task are not described in this paper because they are not relevant to the questions under investigation. The Memory task was presented to each participant after he or she had completed one of the Linguistic or Nonlinguistic tasks described in the text, and thus it did not interfere with collection of the data reported here.

[5] Participants overall directed a considerable proportion of their attention to the head and upper torsos of agents (0.28, no effect of age or language group) and to theme objects (0.22). This pattern of eyegaze is similar to that reported in previous studies on attention to motion events (e.g., Papafragou et al., 2008).

[6] Children and Greek speakers (both $p<0.05$) were most likely to produce "Neither" responses, which is consistent with the pattern of subevent mention described in the text.

[7] These results hold even if we include events for which adults mentioned neither of the causative subevents in the analysis. Multilevel mixed modeling showed no reliable effects of Language

(English, Greek) or Subevent mention (Both, Neither) on looks to Means or Result regions for any analysis windows. In general, then, the instruction to describe the events seems to have had comparable effects on English- and Greek-speaking adults' attention to Means and Result regions, regardless of whether they mentioned both of the subevents or gave a different (e.g., more general) event description. Despite these similarities, however, it is not clear how eyegaze patterns from trials on which neither subevent was mentioned reflect information gathering in preparation for a particular pattern of event description. We therefore continue to focus on trials in which both subevents were mentioned in the remainder of this section.

[8]Preschoolers in this study were more likely than adults to use deictics (e.g., *here*, *there*) to refer to path endpoints when describing Results (cf. example 9b), perhaps to avoid or delay processing costs associated with accessing and producing contentful noun phrases. Future work might investigate whether differences in access to lexical items contributes to differences in attention to the relevant visual elements.

References

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*, 16–48. doi: 10.106/j.cognition.2005.12.006

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. doi: 10.1016/j.jml.2007.12.005

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

Behrend, D. (1990). The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Development*, *61*, 681–696. doi: 10.1111/j.1467-8624.1990.tb02811.x

Behrend, D., Harris, L., & Cartwright, K. (1995). Morphological cues to verb meaning: Verb inflections and the initial mapping of verb meanings. *Journal of Child Language, 22,* 89–106. doi: 10.1017/S0305000900009648

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*, 653–685. doi: 10.1016/S0749-596X(03)00007-X

Bock, K., Irwin, D. E., Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds), *The interface of language, vision, and action* (pp. 249–278). New York: Psychology Press.

Bowerman, M. (1989). Learning a semantic system: What role do cognitive predispositions
    play? In M. L. Rice & R. L. Schiefelbusch (Eds.), *The teachability of language* (pp. 133–
    169). Baltimore: Paul H. Brookes Publishing Co.

Bunger, A. (2006). How we learn to talk about events: Linguistic and conceptual constraints on
    verb learning. Unpublished Ph.D. dissertation, Northwestern University.

Bunger, A., Baier, R., & Lidz, J. (2009). *Two-year-olds think that results lie at the core of
    causative events*. Poster presented at the Biennial Meeting of the Society for Research in
    Child Development. Denver, CO, April 2–4, 2009.

Bunger, A., Papafragou, A., & Trueswell, J. (2013). Event structure influences language
    production: Evidence from structural priming in motion event description. *Journal of
    Memory and Language, 69*, 299-323. doi:10.1016/j.jml.2013.04.002

Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension
    and utterance formulation in children: Evidence from linguistic omissions. *Cognition*, *122*,
    135–149. doi: 10.1016/j.cogntion.2011.10.002

Casasola, M., & Cohen, L. B. (2000). Infants' association of linguistic labels with causal actions.
    Developmental Psychology, *36*, 155–168. doi: 10.1037/0012-1649.36.2.155

Cohen, L. B., Rundell, L. J., Spellman, B. A., & Cashon, C. H. (1999). Infants' perception of
    causal chains. *Psychological Science*, *10*, 412–418. doi: 10.1111/1467-9280.00178

Dessalegn, B., & Landau, B. (2013). Interaction between language and vision: It's momentary,
    abstract, and it develops. *Cognition, 127*, 331-344. doi:10.1016/j.cognition.2013.02.003

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta
    Psychologica*, *12*, 129–143. doi: 10.1016/j.actpsy.2006.07.004

Evans, N., Gaby, A., Levinson, S.C., & Majid, A., Eds. (2011). *Reciprocals and Semantic Typology.* Amsterdam: Benjamins.

Forbes, J. N., & Farrar, M. J. (1993). Children's initial assumptions about the meaning of novel motion verbs: Biased and conservative? *Cognitive Development*, *8*, 273–290. doi: 10.1016/S0885-2014(93)80002-B

Forbes, J. N., & Farrar, M. J. (1995). Learning to represent word meaning: What initial training events reveal about children's developing action verb concepts. *Cognitive Development*, *10*, 1–20. doi: 10.1016/0885-2014(95)90016-0

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, *49*, 988–998.

Gentner, D., & Goldin-Meadow, S., Eds. (2003). *Language and mind.* Cambridge, MA: MIT Press.

Giannakidou, A., & Merchant, J. (1999). Why Giannis can't scrub his plate clean: On the absence of resultative secondary predication in Greek. In A. Mozer (Ed.), *Greek Linguistics '97: Proceedings of the 3rd International Conference on Greek Linguistics* (pp. 93–103). Athens: Ellinika Grammata.

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*, 544–569. doi: 10.1016/j.jml.2007.01.007

Gleitman, L., & Papafragou, A. (in press). New perspectives on language and thought. In K. Holyoak and R. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning*. New York: Oxford University Press.

Göksun, T., George, N., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Forces and motion: Causal understanding of preschoolers. *Child Development*, *84*, 1285–1296. doi: 10.1111/cdev.12035

Göksun, T., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Konishi, H., & Okada, H. (2011). Who is crossing where? Infants' discrimination of figures and grounds in events. *Cognition*, *121*, 176–196. doi: 10.106/j.cognition.2011.07.002

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279. doi: 10.1111/1467-9280.00255

Griffin, Z. M., & Oppenheimer, D. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 943–948. doi: 10.1037/0278-7393.32.4.943

Griffin, Z. M., & Spieler, D. H. (2006). Observing the what and when of language production for different age groups by monitoring speakers' eye movements. *Brain and Language*, *99*, 272–288. doi: 10.1016/j.bandl.2005.08.003

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, *41*, 153–196. doi: 10.1016/0010-0277(91)90035-3

Hafri, A., Papafragou, A., & Trueswell, J. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*, 880–905 doi: 10.1037/a0030045

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

Ji, Y., Hendriks, H., & Hickmann, M. (2011). How children express caused motion events in Chinese and English: Universal and language-specific influences. *Lingua*, *121*, 1796–1819. doi: 10.1016/j.lingua.2011.07.001

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A.
(2010). English speakers attend more strongly than Spanish speakers to manner of motion
when classifying novel objects and events. *Journal of Experimental Psychology: General,
139*, 638–653. doi: 10.1037/a0020507

Klettke, B., & Wolff, P. (2003). Differences in how English and German speakers talk and
reason about CAUSE. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual
Conference of the Cognitive Science Society*, Mahwah, New Jersey: Lawrence Erlbaum.

Kopecka, A., & Narasimhan, B., Eds. (2012). *Events of putting and taking: A crosslinguistic
perspective.* Amsterdam: Benjamins.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied
motion. *Journal of Cognitive Neuroscience*, *12*, 48–55. doi: 10.1162/08989290051137594

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants.
Perception, *13*, 287–305. doi: 10.1068/p130287

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

Levelt, W. (1989). *Speaking*. Cambridge, MA: MIT Press.

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

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics
interface.* Cambridge, MA: MIT Press.

Levinson, S. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In
P. Bloom, M. Peterson, L. Nadel and M. Garrett (eds.), *Language and space*, 109-170.
Cambridge, MA: MIT Press.

MacDonald, M. (2013). How language production shapes language form and comprehension.

    Frontiers in Psychology, *4*, 1-16. doi: 10.3389/fpsyg.2013.00226

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Haryu, E., Vanegas, S., Okada, H.,

    Davis, B. S. (2010). A developmental shift from similar to language-specific strategies in

    verb acquisition: A comparison of English, Spanish and Japanese. *Cognition*, *114*, 299-319.

    doi:10.1016/j.cognition.2009.10.002

Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday

    events: A study of cutting and breaking. *Cognition*, *109*, 235–250. doi:

    10.1016/j.cognition.2008.08.009

Malt, B. C., Gennari, S., Imai, M., Ameel, E., Tsuda, N., & Majid, A. (2008). Talking about

    walking: Biomechanics and the language of locomotion. *Psychological Science*, *19*, 232–

    240. doi: 10.1111/j.1467-9280.2008.02074

McDaniel, D., McKee, C., & Garrett, M. F. (2010). Children's sentence planning: Syntactic

    correlates of fluency variations. *Journal of Child Language*, *37*, 59–94. doi:

    10.1017/S0305000909009507

Naigles, L. R. (1996). The use of multiple syntactic frames in verb learning via syntactic

    bootstrapping. *Cognition, 58*, 221–251. doi: 10.1016/0010-0277(95)00681-8

Naigles, L. R., Eisenberg, A. R., Kako, E. T., Highter, M., & McGraw, N. (1998). Speaking of

    motion: Verb use in English and Spanish. *Language and Cognitive Processes*, *13*, 521–549.

    doi: 10.1080/016909698386429

Naigles, L. R., & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish:

    Influences of language and syntax. *Psychological Science*, *9*, 363-369. doi: 10.1111/1467-

    9280.00069

Nappa, R., Wessell, A., McEldoon, K. L., Gleitman, L. R., Trueswell, J. C. (2009). Use of

   speaker's gaze and syntax in verb learning. *Language Learning and Development*, *5*, 203–

   234. doi: 10.1080/15475440903167528.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time

   course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*. Advance

   Online Publication. doi:10.1080/23273798.2015.1006238Oakes, L. M., & Cohen, L. B.

   (1990). Infant perception of a causal event. *Cognitive Development*, *5*, 193–207. doi:

   10.1016/0885-2014(90)90026-P

Özyürek, A., Allen, S., Furman, R., Kita, S., Brown, A., & Ishizuka, T. (2008). Development of

   cross-linguistic variation in speech and gesture: Motion events in English and Turkish.

   Developmental Psychology, *44*, 1040–1054. doi: 10.1037/0012-1649.44.4.1040

Papafragou, A., Hulbert, J., & Trueswell, J. C. (2008). Does language guide event perception?

   Evidence from eye movements. *Cognition*, *108*, 155–184. doi:

   10.1016/j.cognition.2008.02.007

Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek

   presupposes: The cross-linguistic encoding of motion events. *Cognition*, *98*, B75–B87. doi:

   10.1016/j.cognition.2005.05.005

Papafragou, A., & Selimis, S. (2010). Lexical and structural biases in the acquisition of motion

   verbs. *Language Learning and Development*, *6*, 87–115. doi: 10.1080/15475440903352781

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*, 47–81. doi: 10.1016/0010-

   0277(91)90032-Y

Russo, J. E., & Rosen, L. D. (1975). An eye fixation analysis of multialternative choice. *Memory

   and Cognition*, *3*, 267–276. doi: 10.3758/BF03212910

Scerif, G., Karmiloff-Smith, A., Campos, R., Elsabbagh, M., Driver, J., & Cornish, K. (2005). To

    look or not to look? Typical and atypical development of oculomotor control. *Journal of*

    Cognitive Neuroscience, *17*, 591–604.

Shipley, T., & Zacks, J., Eds. (2008). *Understanding events: From perception to action.* New

    York: Oxford University Press.

Skordos, D., & Papafragou, A. (2014). Lexical, syntactic and semantic-geometric factors in the

    acquisition of motion predicates. *Developmental Psychology, 50,* 1985-1998.

Slobin, D. I. (1985). Cross-linguistic evidence for the language-making capacity. In D. I. Slobin

    (Ed.), *The cross-linguistic study of language acquisition, Vol. II: Theoretical issues* (pp.

    1157–1256). Hillsdale, NJ: Erlbaum.

Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking'. In J. Gumperz & S.

    Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). New York: Cambridge

    University Press.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression

    of motion events. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative: Vol.*

    2. Typological and contextual perspectives (pp. 219–257). Mahwah, NJ: Lawrence Erlbaum.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen

    (Ed.), *Grammatical categories and the lexicon* (Vol. III, pp. 57–149). Cambridge: Cambridge

    University Press.

Talmy, L. (2003). *Toward a cognitive semantics: Typology and process in concept structuring.*

    Cambridge, MA: MIT Press.

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-

linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, *63*,

64–82. doi: 10.1016/j.jml.2010.02.006

Wilson, F., Unal, E., Trueswell, J., & Papafragou, A. (in prep). Homologies between language

and event cognition. Paper to be published in *Proceedings of the 39th annual Boston*

University Conference on Language Development. Somerville, MA: Cascadilla Press.

Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transtive action

observation have sequential structure. *Acta Psychologica*, *133*, 51–56. doi:

10.1016/j.actpsy.2009.09.001

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. T. (2011). Rapid extraction of event

participants in caused motion events. In L. Carlson, C. Hölscher & T. Shipley (Eds.),

Proceedings from the 33rd Annual conference of the Cognitive Science Society (pp. 1206–

1211). Hillsdale, NJ: Erlbaum.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*,

82–111. doi: 10.1037/0096-3445.136.1.82

Wolff, P. (2008). Dynamics and the perception of causal events. In T. Shipley & J. Zacks (Eds.),

Understanding events: How humans see, represent, and act on events (pp. 555–587). New

York: Oxford University Press.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews:*

Cognitive Science, 2, 253-265. doi: 10.1002/wcs.104

Wolff, P., Jeon, G., Klettke, B., & Yu, L. (2010). Force creation and possible causers across

languages. In B. Malt, & P. Wolff (Eds.), *Words and the mind: How words capture human*

experience. Oxford: Oxford University Press.

Wolff, P., Jeon, G., & Yu, L. (2009). Causers in English, Korean and Chinese and the

   individuation of events. *Language and Cognition*, *1,* 167-196. doi:

   10.1515/LANGCOG.2009.009

Wolff, P., & Ventura, T. (2009). When Russians learn English: How the semantics of causation

   may change. *Bilingualism: Language and Cognition, 12,* 153-176. doi:

   10.1017/S1366728909004040

Zheng, M., & Goldin-Meadow, S. (2002). Thought before language: How deaf and hearing

   children express motion events across cultures. *Cognition*, *85*, 145–175. doi: 10.1016/S0010-

   0277(02)00105-1

Appendix

Table A1

Causative events depicted in target images, sample words used to describe Means and Result

subevents, and content of Means and Result regions used in the eyegaze analysis.

Target event (Means; Result)	Subevents	Sample words used in descriptions (English; Greek)	Image region (Size in deg)
A man hits a basketball using his arms; the basketball flies over a volleyball net.	Means	hit; n/a*	the man's arms and lower body (11.0 x 8.3)
	Result	over; se (*at*)	the net and space beyond it in the potential path of the ball (11.9 x 7.0)
A mouse pushes a wedge of cheese with its arms; the cheese goes into the mouse's hole.	Means	push; sprohni (*push*)	the mouse's arms and lower body (11.2 x 6.6)
	Result	into; mesa stin (*into*)	the doorway (5.3 x 16.1)
A sheep kicks a beach ball with its hind legs; the ball flies over/into a box.	Means	kick; htipai (*hit*)	the sheep's lower body (7.0 x 5.4)
	Result	into; mesa (*inside*)	the box and space above it in the potential path of the ball (composed of 2 AOIs added together: 13.6 x 8.1, 7.6 x 9.7)
A boy squeezes a bottle of mustard with his hand; the mustard falls onto on a hot dog.	Means	squirt; rihni (*throw*)	the boy's hand and the bottle (4.0 x 5.9)
	Result	onto; sto (*at*)	the hot dog (8.3 x 6.2)
A boy hits a soccer ball with his fist; the ball flies (in)to a basket.	Means	punch; htipai (*hit*)	the boy's fist (4.0 x 3.9)
	Result	into; sto (*at*/*into*)	the basket (9.7 x 14.0)
A woman lifts the back of her car with her	Means	push; vali (*put*)	the woman's arms and lower body (6.2 x 5.7)

hands and pushes the car forward; the car rolls (in)to a garage.	Result	into; mesa sto (*into*)	the garage (10.7 x 17.5)
A man throws a rugby ball with his arms; the ball flies into a hockey net.	Means	throw; petai (*throw*)	the man's arms and torso (10.2 x 4.4)
	Result	into; mesa se (*into*)	the hockey net (11.8 x 16.8)
A girl throws a vase with her hands; the vase falls into a basket.	Means	throw; petai (*throw*)	the girl's arms and hands (5.9 x 2.3)
	Result	into; mesa se (*into*)	the basket (8.1 x 8.1)
A woman sweeps a rake toward a leaf; the leaf goes (in)to a shopping bag.	Means	rake; mazevi (*gather*)	the rake and the woman's torso and arms (composed of 3 AOIs added together: 4.3 x 6.8, 1.7 x 1.9, 6.0 x 3.7)
	Result	into; mesa se (*into*)	the bag, the leaf, and intervening space in the potential path of the leaf (15.7 x 5.1)
A man holding an axe drags a cut pine tree; the man and tree move (in)to a tent.	excluded		
A large hand flicks a snail; the snail moves into a basket.	excluded		

Region sizes are given for images shown to adults in degrees of visual angle at a viewing distance of 60 cm. Preschoolers viewed the videos under slightly different display settings, resulting in an approximately 50% decrease in the size of the regions for preschoolers as compared to adults. [*]Although there are verbs available in Greek that could have been used to describe the Means of the first target event listed in the Appendix (e.g., xtipai (*hit*) or rixni (*throw*)), no Greek-speaking participants mentioned the Means subevent in their description of it.

Instead, they tended to give a more general description of the event, e.g., enas kirios pezi volei

"A man playing volleyball."

Table 1

Proportion of mention of Means and Result subevents in descriptions of target causative events

	Both subevents	Means subevent only	Result subevent only
All participants	0.47 (± 0.05)	0.30 (± 0.05)	0.24 (± 0.03)
English			
Adults	0.79 (± 0.05)	0.08 (± 0.04)	0.13 (± 0.02)
4-year-olds	0.29 (± 0.06)	0.40 (± 0.08)	0.31 (± 0.06)
3-year-olds	0.13 (± 0.04)	0.66 (± 0.10)	0.21 (± 0.08)
Greek			
Adults	0.62 (± 0.09)	0.12 (± 0.06)	0.27 (± 0.06)
4-year-olds	0.11 (± 0.07)	0.52 (± 0.11)	0.37 (± 0.10)
3-year-olds	0.19 (± 0.10)	0.50 (± 0.13)	0.31 (± 0.15)

Values represent the proportion of use of each information packaging type (± standard error) and

are taken just from event descriptions that include mention of one or more causative subevents.

Table 2

Fixed effects from best-fitting multilevel linear model of number of subevents mentioned

Effect	Estimate	S.E.	t-value
Intercept	1.21	0.07	17.41
Age: 3yo vs. adult Language: English vs. Greek	0.53	0.08	-6.23
Age: 3yo vs. 4yo	0.09	0.08	1.11
Language: English vs. Greek	-0.19	0.07	-2.58

Formula in R: DepVar ~ Language + Age + (1 | Subject) + (1 + Language × Age | Item).

Table 3

Fixed effects from best-fitting multilevel linear model of single subevent mentioned

Effect	Estimate	S.E.	t-value
Intercept	0.26	0.11	2.39
Age: 3yo vs. 4yo	0.14	0.08	1.75
Age: 3yo vs. adult	0.32	0.08	3.87

Formula in R: DepVar ~ Age + (1 | Subject) + (1 + Age | Item).

Table 4

Fixed effects from best-fitting multilevel linear models of Means and Result looking by time

window, adults

Effect	Estimate	S.E.	t-value
Means looking: 0–1 s			
Intercept	-2.76	0.31	-8.82
Language: English vs. Greek	-0.72	0.25	-2.93
Result looking: 1–2 s			
Intercept	-3.76	0.33	-11.26
Task: Linguistic vs. Nonlinguistic	1.39	0.39	3.52
Result looking: 3–4 s			
Intercept	-1.47	0.47	-3.12
Task: Linguistic vs. Nonlinguistic	-0.81	0.39	-2.16

The models presented are the best fitting models for each time window; when effects or

interactions do not appear, it is because adding them to the models did not reliably improve the

fit. Formulas in R: Means looking, 0–1 s: DepVar ~ Language + (1 | Subject) + (1 + Language |

Item); Result looking, 1–2 s and 3–4 s: DepVar ~ Task + (1 | Subject) + (1 + Task | Item).

Table 5

Fixed effects from best-fitting multilevel linear models of Means and Result looking in the

Linguistic Task, preschoolers

Effect	Estimate	S.E.	t-value
Means looking: 0–1 s			
Intercept	-1.84	0.48	-3.81***
Production Type: Means Only vs. Result	-0.79	0.36	-2.19*
Means looking: 1–2 s			
Intercept	-2.44	0.41	-5.90***
Language: English vs. Greek	1.04	0.36	-2.90**
Production Type: Means Only vs. Result	-1.02	0.39	-2.60*
Result looking: 2–3 s			
Intercept	-3.60	0.49	-7.33***
Production Type: Means Only vs. Result	1.29	0.49	2.63**

Formulas in R for best-fitting models in each analysis window:

Means 0–1 s: DepVar ~ Production Type + (1 | Subject) + (1 | Item)

Means 1–2 s: DepVar ~ Language + Production Type + (1 | Subject) + (1 | Item)

Result 2–3 s: DepVar ~ Production Type + (1 | Subject) + (1 | Item)

When factors and interactions do not appear in reported models, it is because including them did

not reliably improve the fit compared to an empty model with no fixed effects.

*p<0.05; **p<0.01; ***p<0.001

Table 6

Fixed effects from best-fitting multilevel linear model of Result looking in the 2–3 s time window,

preschoolers

Effect	Estimate	S.E.	t-value
Result looking: 2–3 s			
Intercept	-2.94	0.35	-8.38**
Task: Linguistic-Result vs. Nonlinguistic	-0.62	0.27	-2.28*

Formula in R: DepVar ~ Condition + (1 | Subject) + (1 | Item). Factors and interactions that do

not appear in the reported model did not reliably improve the fit compared to an empty model

with no fixed effects. *p<0.05; **p<0.001

Figures
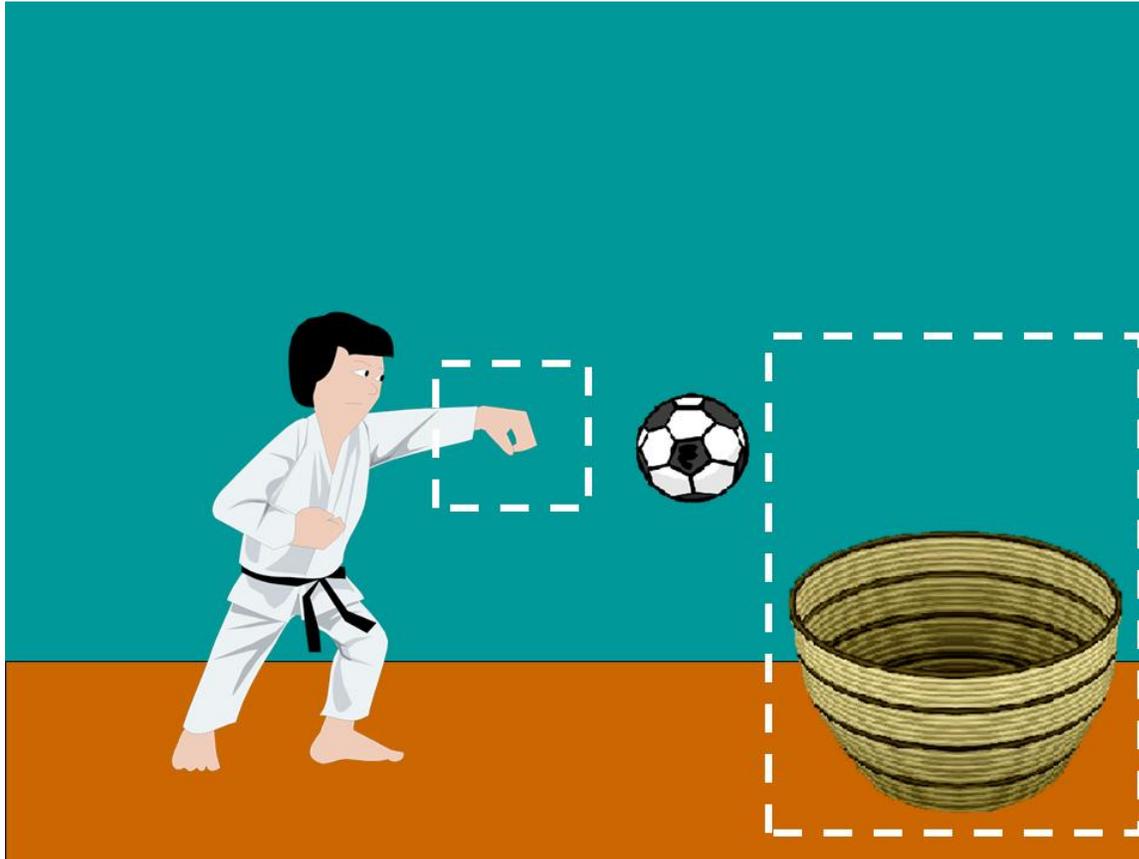
Figure 1. A sample target event. The Means subevent of this complex causative event is the

boy's punching of the ball, and the Result subevent is the ball's motion toward the basket. Dotted

lines indicating Means and Result Regions were not visible to study participants.
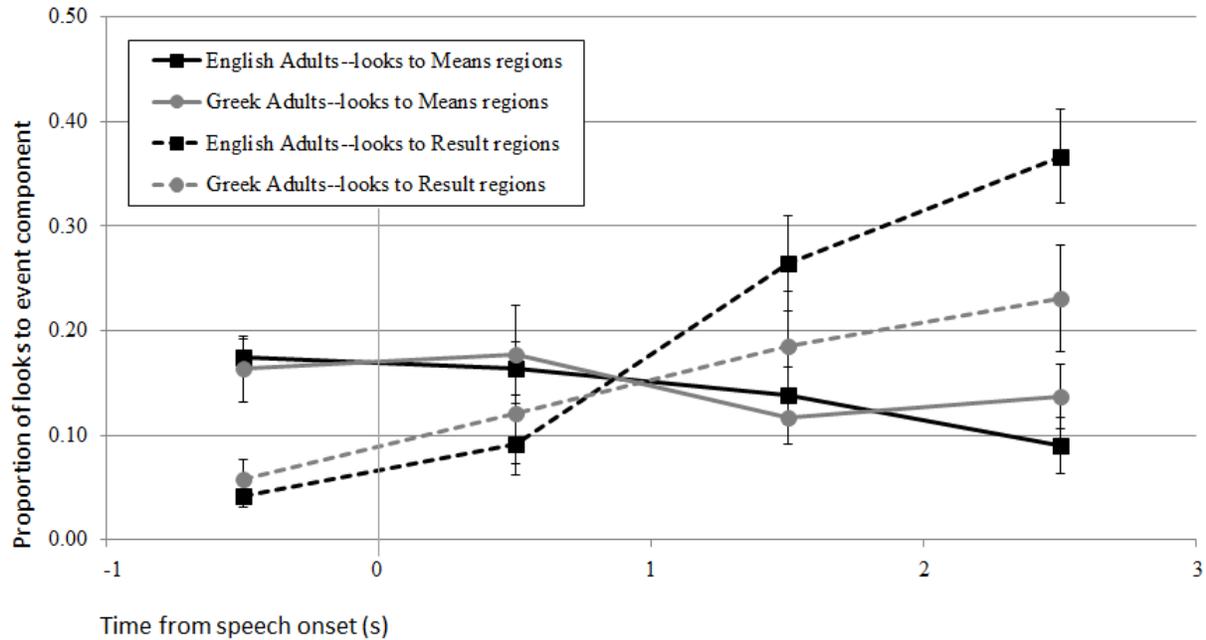
Figure 2. Mean proportion of looks to Means and Result regions by adults in the Linguistic task who mentioned both causative subevents. Data are aligned by speech onset, designated by the vertical line at 0 s. Proportions are calculated based on looks to the entire target image, and so are not expected to add up to 1 across these two regions. No reliable effects of language background on looks to Means or Result regions was found in the analysis.
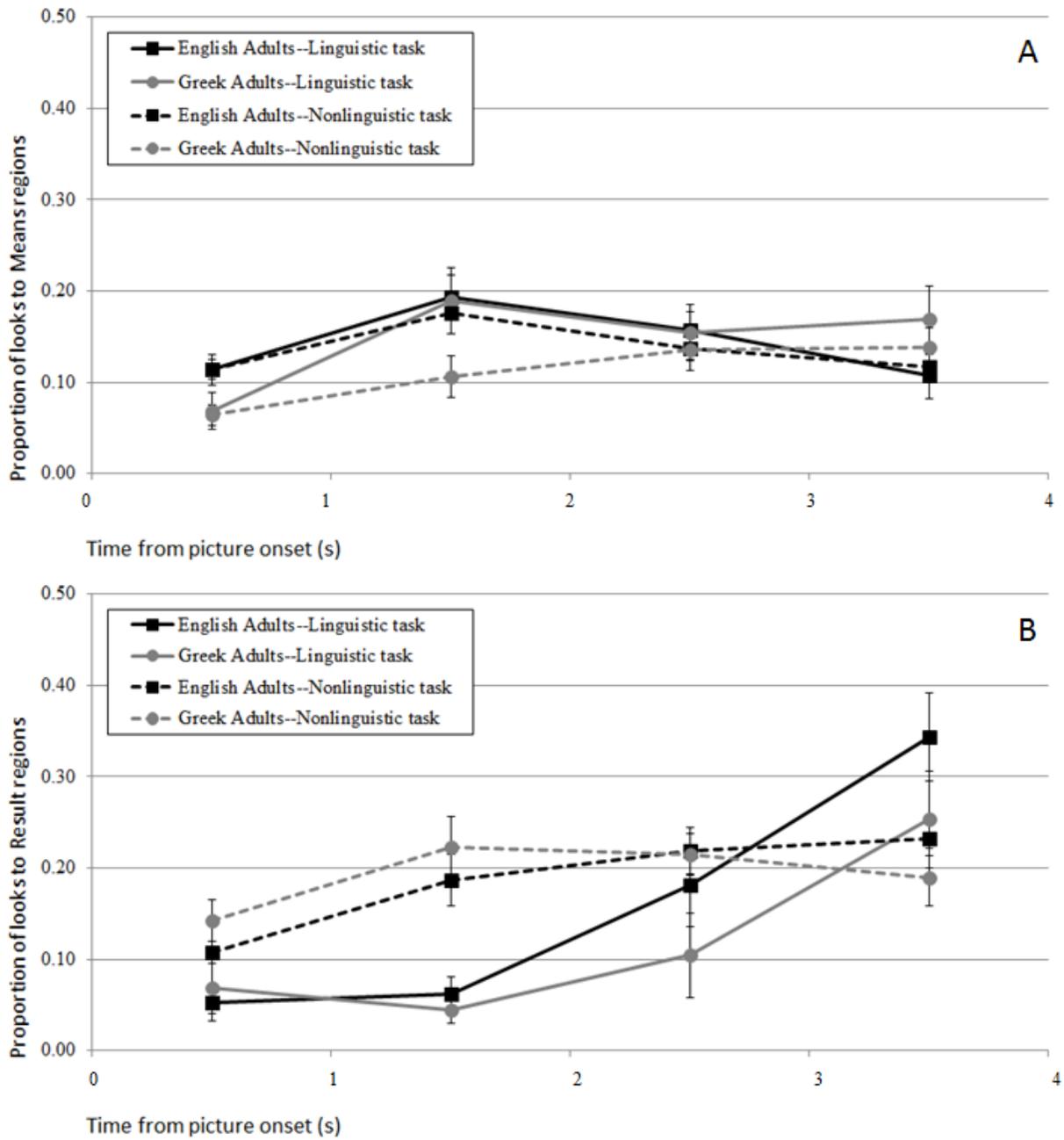
Figure 3. Proportion of looks to Means (A) and Result (B) regions of causative target events by

English and Greek-speaking adults in the Linguistic and Nonlinguistic tasks. Proportions are

calculated based on looks to the entire target image. Data for the Linguistic task are taken from only those trials on which adults mentioned both causative subevents of the target event. Collapsing across tasks, English-speaking adults directed more attention to Means regions (A) during the first second after the onset of the stimulus image than Greek adults did ($p < 0.05$). Across language groups, task was a reliable predictor of looks to Result regions (B) in seconds 1–2 ($p < 0.01$) and 3–4 ($p < 0.05$).
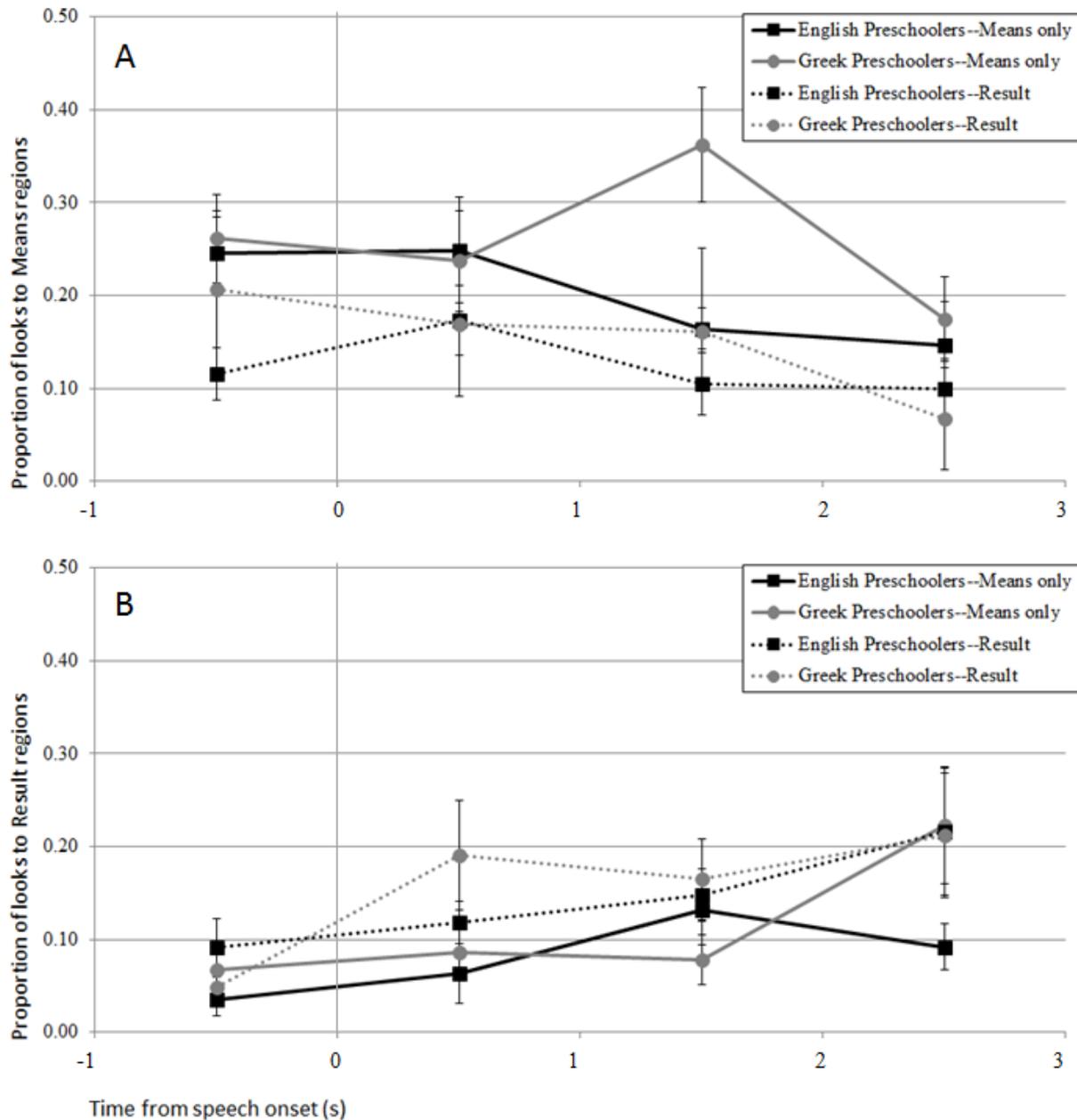
Figure 4. Mean proportion of looks to Means (A) and Result (B) regions by preschoolers in the Linguistic task. Data are aligned by speech onset, designated by the vertical line at 0 s, and are split by whether just the Mean subevent or the Result subevent (with or without the Means) was mentioned in event descriptions. Proportions are calculated based on looks to the entire target image, and so are not expected to add up to 1 across these two regions. Collapsing across

languages, preschoolers reliably directed more attention to Means regions (A) during seconds 0–2 for trials on which they mentioned just Means information (p<0.05) and more attention to Result regions (B) during second 2–3 of trials on which they mentioned Results (p<0.01). Collapsing across utterance content, English-speaking preschoolers directed more attention to Means regions (A) during second 1–2 than Greek-speaking preschoolers did (p<0.01).
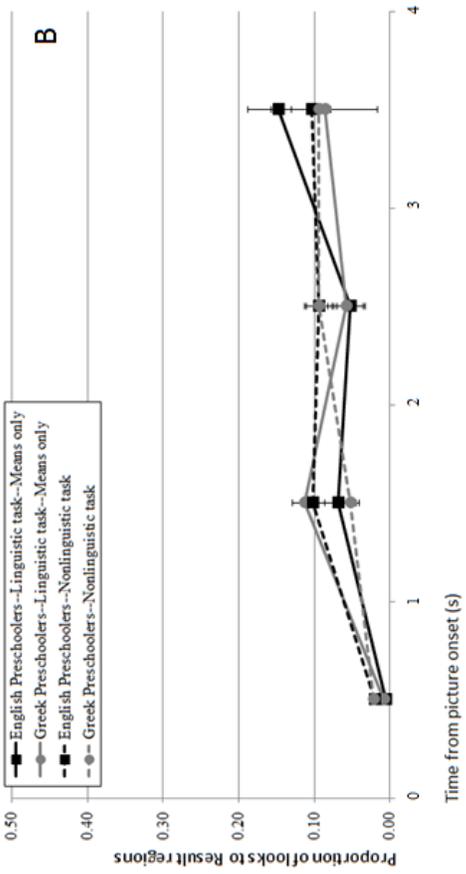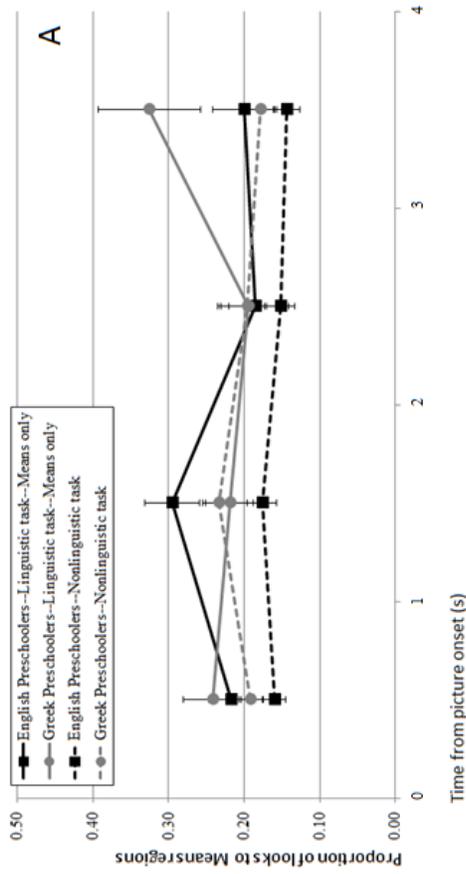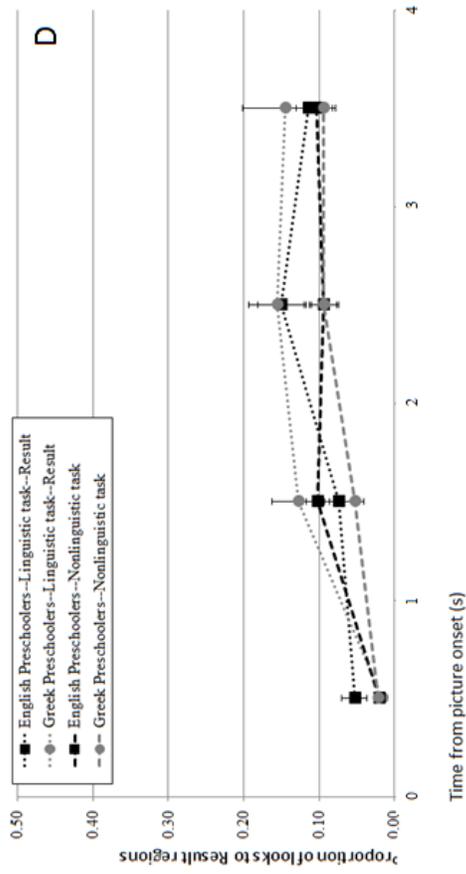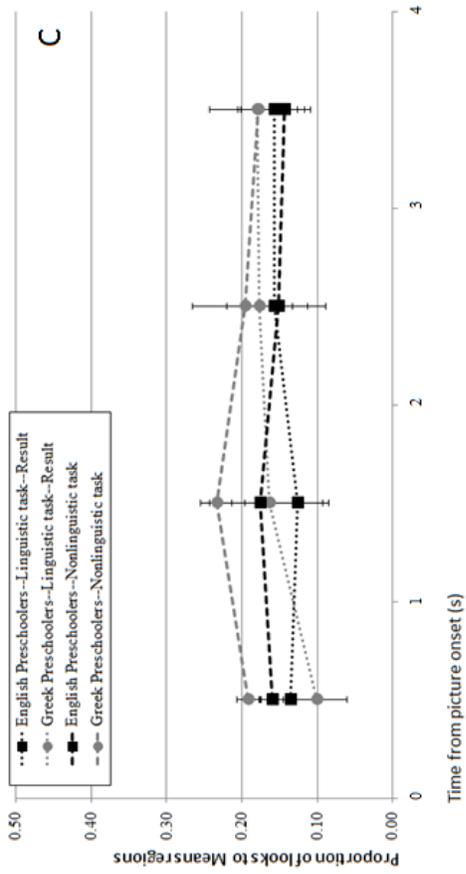
Figure 5. Proportion of looks to Means (A,C) and Result (B,D) regions of causative target events by English and Greek-speaking preschoolers in the Linguistic and Nonlinguistic tasks. Data from the Linguistic task are split by whether just the Means subevent (A,B) or the Result subevent (with or without the Means, C,D) was mentioned in event descriptions. Data from the Nonlinguistic task are repeated across figures A and C and figures B and D. Proportions are calculated based on looks to the entire target image. Task was a reliable predictor of looks to Result regions when participants mentioned Result information (D): collapsing across language groups, preschoolers who mentioned Results directed more attention to Result regions during second 2–3 than did preschoolers in the Nonlinguistic task ($p<0.05$). There were no other reliable effects of Language or Task or interactions between the two.