Seeing and Saying: The Relation Between Event Apprehension and Utterance Formulation in Children

Ann Bunger[1], John Trueswell[2], and Anna Papafragou[1]
[1]University of Delaware and [2]University of Pennsylvania

1. Introduction

In adults, there is a clear link between attention allocation and event conceptualization during language planning, such that allocation patterns during speech planning are predictive of the structure and content of linguistic output. When viewing an event while preparing to describe what they see, adults very quickly direct their attention to components of the event that they plan to talk about, and usually in the order that they plan to mention them. For example, when asked to describe a still image of an event in which a mouse sprays a turtle with a water gun, adults fixate event participants about 1 sec *before* they mention them, turning their attention first to the mouse and then to the turtle when planning a description like "The mouse is spraying the turtle" (Griffin & Bock, 2000; see also Bock, Irwin & Davidson, 2004; cf. Gleitman, January, Nappa & Trueswell, 2007 and Papafragou, Hulbert & Trueswell, 2008 for similar findings during the viewing of dynamic events). This clear direction of attention to select event components over time disappears when language is not involved in the task of scene inspection or when access to language is blocked during event viewing, (Bock et al., 2004; Griffin & Bock, 2000; Papafragou et al., 2008). In the current study, we ask whether this same relationship between event apprehension and language planning exists in children. We also ask whether children's attention allocation during speech planning reflects their linguistic output, which is typically less detailed than adults', or a richer, more adult-like event conceptualization.

There is a long tradition in psycholinguistics of looking at what children say to get information about what they know—that is, of using the sophistication of their utterances to estimate their conceptual sophistication (e.g., Slobin, 1973). One very obvious fact that is commonly observed about children's utterances is that they are not as rich as those of adults. Children's utterances tend to be shorter than those of adults, and mean length of utterance (MLU) continues to grow throughout the early school years: from around 2 words per utterance at 2 years of age to 5 words per utterance in 7-year-olds and up to 6.6 words per

* This research was supported in part by NIH/NICHD Grant #1-R01-HD055498 to A. Papafragou and J. Trueswell.

utterance in adults (Brown, 1973; Rice, Redmond & Hoffman, 2006; Weijer, 1999). In addition, children's utterances often lack the grammatical complexity of adult utterances: e.g., children do not make use of the full range of structures that are available to adults and they often omit words in ways that are not adult-like.

These observed differences in production are often used as support for the claim that children have impoverished conceptual systems in comparison to adults (e.g., Bartsch & Wellman, 1995; Gentner, 1982; Huttenlocher, Smiley & Charney, 1983; Johnston & Slobin, 1978; Perner, Sprung, Zauner & Haider, 2003; Piaget, 1955; Poulin-Dubois & Graham, 2007). Roughly, the argument is that if children lack the word or the grammatical structure that is used by adults to express a particular concept, this is taken as evidence that children lack the concept itself. Johnston & Slobin (1978) provide an example of this kind of account for the acquisition of spatial terms, arguing that the conceptual complexity of various spatial expressions determines their order of acquisition, such that words that label simpler concepts are acquired before those that label more complex concepts. On the basis of the finding that across several languages, expressions meaning something like "in" were used by children before expressions meaning something like "between," Johnston and Slobin argue that working out the meaning of a containment expression like "in" is easier because it requires calculating a relationship between just two objects—some object and the container it goes into—compared to three objects for a between relationship. Recent work in event categorization suggests that there is some truth in this conceptual hierarchy: children do, in fact, form abstract categorizations of containment events (at 6 mo; Casasola, Cohen & Chiarello, 2003; Quinn, 2005) before "between" events (at 10 mo; Quinn, 2005).

Accounts of this nature, which question the developmental continuity of conceptual capacity, can be contrasted with those in which differences in children's utterances compared to adults are due to constraints on language processing or language learning (e.g., L. Bloom, 1970; P. Bloom, 1990; Pinker, 1984; Slobin, 1973). According to such accounts, children may not be producing certain linguistic elements not because of a lack of conceptual sophistication, but rather because of limits on production, which could place restrictions on the length of utterances. As a result of these restrictions, certain elements that children intend to include in their message may fail to surface. A very basic example of this kind of argument comes from the syntactic bootstrapping literature, in which it has been demonstrated that children as young as 2 years of age can *understand* complex, multi-word sentences even though they might not yet be able to *produce* utterances of more than one or two words in length (Fisher, 1996, 2002; Gleitman, 1990; Naigles, 1990; inter alia).

In the current study, we investigate whether reported differences in production between child and adult populations are really a sign that children lack conceptual sophistication. To accomplish this, we focus on children's descriptions of simple motion events in which an animate agent moves in a particular manner along a particular path. Although motion events are of interest

crosslinguistically, we discuss data only from English speakers. When adult speakers of English describe motion events, they are likely to encode the manner of motion in the main verb of the sentence and to include path information, if it is given at all, in a satellite of the verb (e.g., Slobin, 1996; Talmy, 1985). Recent work also suggests that 5-year-old English-speaking children are likely to follow this pattern when describing motion events (Papafragou & Selimis, 2010; Slobin, 1996). In the study described here, we look more closely at the nature of children's descriptions of motion events and compare them to what children are looking at while they watch them.

2. Methods

2.1 Participants

Data were collected from 20 5-year-old children (age 4;7, range 4;1–5;1) and 20 adults. Children were recruited from preschools in Newark, DE and Philadelphia, PA; adults were students at the University of Delaware or the University of Pennsylvania and received course credit for participation. All study participants are native monolingual speakers of English.

2.2 Materials

The stimuli consisted of silent events that had been created by animating clip-art scenes. Target events depicted 12 simple motion events in which an animate agent moved in a particular manner to reach a stationary path endpoint. Manners and Paths in our events are represented by distinct visual elements: each agent uses an instrument that determines its Manner of motion, and they always move toward a visible endpoint that determines the Path. A still frame from one of the target events is given in Fig. 1: in this event a boy (the agent) moves on rollerskates (the instrument) across the screen and into a soccer net (the path endpoint). Clipart images were constructed such that the instrument was spatially separated from the torso and face of the agent. Filler events presented animate agents involved in dynamic events that did not include a specific endpoint (e.g., flying a kite). For all events, the animated movement lasted for just 3 sec, at which time a beep was heard. The final frame of the event remained visible for 6 sec more for preschoolers and 2 sec more for adults, giving participants time to study the scene and to complete assigned tasks.

Two modified versions of each target video were created for use in the memory task. One version involved a change to the manner of motion used by the agent (Manner change, e.g., substituting a skateboard for the rollerskates). The other version involved a change to the path endpoint (Path change, e.g., substituting a gazebo for the soccer net). Path changes always resulted in linguistically relevant differences in the relation between the agent and the path landmark: e.g., instead of skating *into* the soccer net, the boy skates *past* the gazebo.

Figure 1—Still frame taken from a target event

2.3 Procedure

Participants viewed a sequence (one of two fixed semi-randomized lists) of the 24 target and filler events and performed one of two randomly-assigned tasks while watching these stimuli. Half of the participants in each age group were asked to watch each event closely in preparation for a memory game (Nonlinguistic Task), and the other half were asked to provide a verbal description of the event when the beep sounded (Linguistic Task). Immediately after viewing the full set of events, all participants completed a memory task. For this task, they were asked to watch modified versions of the 24 stimulus events and to judge whether each event was the same or different from the versions they had seen before. All of the filler events shown during the memory task were identical to those that participants had seen before, and all of the target events were different—half had changes to the manner of motion used in the event and half had changes to the path endpoint. Each study participant saw only one kind of change for each target event—either a Manner change or a Path change. Manner and Path changes were distributed across the two lists so that an event that appeared with a Manner change in one list appeared with a Path change in the other. Presentation of events during the memory task was experimenter-paced, so that participants could move on to a new event as soon as a judgment had been provided for the previous one.

Stimuli were presented to all participants on the same Tobii 1750 remote eyetracking system. The data sampling rate was 50 Hz, and spatial resolution of the tracker was approximately 0.5 degrees of visual angle. Sessions in which participants completed the Linguistic task were audio recorded to facilitate

accurate collection of event descriptions. Experimenters recorded judgments made during the memory task by hand.

2.4 Data analysis

Event descriptions collected from participants in the Linguistic condition were transcribed and coded by hand to assess the event components that speakers in each age group encoded in descriptions of motion events. Words or phrases that referred to instruments depicted in the target events (e.g., "rollerskates") or ways in which they facilitated the agent's motion (e.g., "skater," "skating," "riding") were coded as mentions of the Manner of motion, and those that referred to the path endpoint (e.g., "goal," "hockey net") or the direction of motion (e.g., "across," "into") were coded as Path mentions. Judgments provided during the memory task were analyzed for the effects of changed event component, age, and task on accuracy.

Eye movement data were analyzed to assess the effects of age and task on encoding of event components. Only eye movements from the encoding tasks during target trials were included in the analysis, excluding data collected for fillers and during the memory task. Track loss was determined separately for each eye by Tobii's ClearView software. Gaze coordinates were taken from eyes with no track loss (or from an average of both eyes, if neither eye had track loss). Trials with global track loss of >33% were excluded from the analysis (n=1 for adults in the Nonlinguistic condition; n=4 for preschoolers in the Nonlinguistic condition; n=8 for preschoolers in the Linguistic condition). Subjects with four or more excluded target trials were excluded from the eye movement analysis (n=1 preschooler in the Linguistic condition).

Dynamic spatial scoring regions were defined for each event around agents, instruments, and path endpoints. An automated analysis procedure allowed these scoring regions to move with the relevant event component during the unfolding of the event. Looks to instruments are interpreted as attention to Manners of motion; looks to path endpoints are interpreted as attention to Paths of motion (see Papafragou et al., 2008, for justification of this link between spatial regions and event components). Overlap between areas of interest (which arose only at the end of events, when agent/instruments overlapped with path endpoints) was resolved by assigning that look to the event component that was visible in the top layer of the visual display, which was always the agent or instrument.

3. Results

3.1 Event descriptions

Not surprisingly, preschoolers produced shorter event descriptions than adults did (MLU in words = 5.6 for preschoolers, 7.6 for adults; p=0.03). Looking more closely at the information content of these utterances allows us to investigate whether age-related differences in utterance length correspond to differences in the linguistic encoding of event components. Table 1 shows the

proportion of event descriptions provided by study participants in which Manner- and Path-related event components were encoded in any linguistic elements (i.e., giving equal weight to components encoded in verb and non-verb elements). Both preschoolers and adults were more likely to mention the Manners of motion events than their Paths. Preschoolers omitted Manners more often than adults did, but what they omitted from their utterances most often was information about the Path of the motion event, including Paths in only 33% of their event descriptions.

Table 1—Distribution of event components in linguistic descriptions

	Manner	Path
Adults	1.00	0.74
Preschoolers	0.82	0.33

Values indicate proportions of event descriptions in which Manner and Path components were encoded in any linguistic elements.

For the rollerskating event depicted in Fig. 1, for example, most adults provided a description that included both the skates (Manner) and the net (Path), as in (1a). Preschoolers on the other hand, were more likely to provide a description like (1b) that included only the skates and not the net.

(1) a. A boy rollerbladed into a soccer net. Adult
b. The boy was skating. Preschooler

Consistent with the overall pattern in their language, both groups of subjects were more likely to encode Manners rather than Paths in verbs. When preschoolers did include Paths in their event descriptions, like adults they were most likely to put them in a satellite of the verb rather than in the verb itself. The descriptions in (2) were provided for a motion event in which an alien drives his car (Manner) into the mouth of a cave (Path).

(2) a. An alien is driving his car into a cave. Adult
b. He was driving inside of a cave. Preschooler

This similarity between the two age groups suggests that the preschoolers' tendency to leave out Path information was probably due to their lower overall utterance length than to a different pattern of event encoding. Because they tended to produce fewer words, children had to leave something out of their utterances, and the thing that they ended up leaving out was the same event component that adult speakers of their language tend to encode in a verbal satellite rather than in the verb itself.

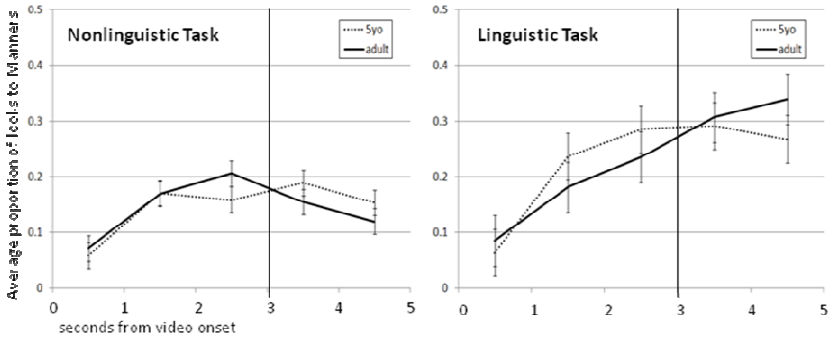
3.2 Eye movement data

Given the finding that preschoolers tended to include Paths in their event descriptions less often than adults did, we turn to the eye movement data to ask whether they were also paying less attention to this category of event components while viewing the events. Gaze patterns during the Nonlinguistic task reveal whether each event component was available to the conceptual system, and gaze patterns during the Linguistic task reveal whether Manner and Path were available to language. If preschoolers omitted Paths from event descriptions because of a conceptual limitation on event encoding, then Path as a conceptual category should not be available to them in either task, and we should not find them attending to Paths in the motion events as much as adults do. If, on the other hand, they omitted Paths because of limits on language processing, then they should pay as much attention to Paths as adults do in both tasks.

Informal inspection of the eye movements depicted in Figure 2 reveals important similarities in allocation of attention to event components across age groups. In both tasks, adults and preschoolers allocate their attention to Manners similarly throughout the viewing period (Fig. 2A). In the Nonlinguistic task, the amount of attention that both age groups allocate to Manners remains fairly steady throughout the viewing period. In the Linguistic task, however, there is a gradual increase in attention to Manners over the viewing period for both age groups. This increase in attention to Manners is consistent with a strategy in which viewers are allocating more attention to the event component they plan to talk about (in English, instruments encoded in manner verbs). Inspection of Figure 2B reveals, moreover, that adults and preschoolers in both tasks are allocating approximately the same amount of attention to Paths and on the same time course. This similarity in allocation of attention to Paths across age groups and tasks suggests that at some level of processing, this event component is just as available to preschoolers as it is to adults. That is, even if Paths do not make it to the surface in preschoolers' linguistic descriptions, they are available for sentence planning.

We tested the reliability of the patterns observed in Figure 2 using multilevel mixed elogit modeling, following Barr (2008). Nonaveraged Subject and Item data were included in the model as crossed random effects (after Baayen, Davidson & Bates, 2008). Path minus Manner preference was modeled in 1-s windows using Task (Linguistic, Nonlinguistic) and Age (Adult, Preschool) as first-level predictors. Each Subject and Item was given a separate random intercept. Task was a reliable predictor of looking patterns ($p<0.01$) starting at 2 sec, reflecting the increase in looks to Manner in the Linguistic task but not the Nonlinguistic task. There were no effects of or interactions with Age.

A: Manner looks



B: Path looks

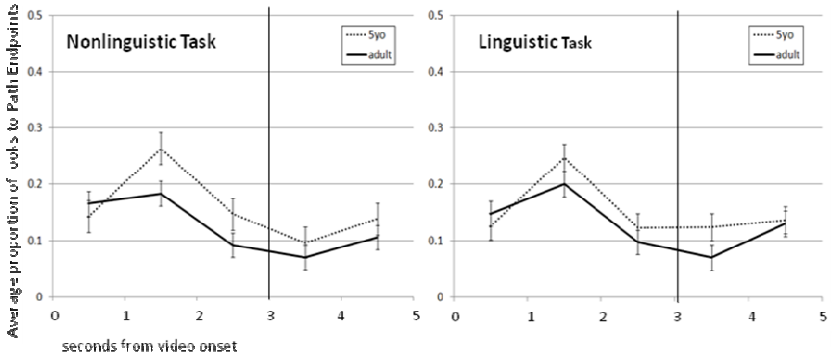


Figure 2—Proportion of looks to Manners (A) and Paths (B) of motion events. Gaze preferences are averaged across 1-sec units of the viewing period. The full viewing period for adults is presented, and just the first 5 sec of viewing for preschoolers. The beep that occurs during the viewing phase happens at the end of second 3.

3.3 Memory data

Analysis of the eyegaze data revealed that preschoolers and adults spent an equivalent amount of time looking at Paths in our motion events. Data from the memory task were assessed to determine whether preschoolers were conceptualizing Paths while they looked at them. If preschoolers notice changes to the Paths of motion events at least as often as they notice changes to Manners of motion, this will provide evidence that they actually encoded Paths as part of their event representations.

Table 2 shows percent accuracy for noticing changes to Manner and Path components during the Memory Task. Across the board, both age groups were

more accurate at picking out changes to Paths than they were at picking out changes to Manners. Even in the Linguistic task, in which preschoolers included information about Path in their descriptions of motion events only a third of the time, 67% of the time they were accurate at identifying changes to Paths. Multilevel hierarchical modeling of percent accuracy was carried out, with Age (Adult, Preschooler), Condition (Nonlinguistic, Linguistic), and Change-Type (Manner, Path) as first-level predictors. Nonaveraged Subject and Item data were included in the model as crossed random effects; each Subject and Item was given a separate random intercept. The best-fitting model included a main effect of Change-Type (p=0.05), with accuracy for Path changes significantly greater than for Manner changes, and an Age x Condition interaction (p=0.01), with preschoolers performing worse on the memory task in the Linguistic condition. This discrepancy in accuracy across conditions can be explained in part by the fact that the memory task was a surprise for participants in the Linguistic condition. Moreover, this finding is consistent with studies that suggest that children may not be able to use language strategically to support memory until they are 6 or 7 years of age (cf. Palmer, 2000; Hitch, Halliday, Schaafstal & Heffernan, 1991). In our study, adults' memory for event components improved with linguistic encoding of the event, but linguistic encoding did not seem to provide the same benefit for preschoolers.

Table 2—Percent accuracy in identifying changes to event components

	Adults		Preschoolers	
Task	Nonlinguistic	Linguistic	Nonlinguistic	Linguistic
Manners	72%	87%	65%	53%
Paths	83%	93%	83%	67%

Accuracy on Path changes was independent of language production. Indeed, adding type of utterance (Path-Mention vs. No-Path-Mention) did not produce a reliably better fit of the accuracy data and resulted in a low non-significant beta coefficient. Both preschoolers and adults in the Linguistic condition were just as accurate for events in which they had mentioned Paths in their event descriptions as for those in which they had not. Taken together, the results of the memory task demonstrate that even though preschoolers were not consistently labeling paths in their event descriptions, they were encoding them as part of their nonlinguistic event representations. If they had not conceptualized event Paths, it is not clear how they could have noticed changes to them.

4. Conclusions

The goal of this study was to investigate whether differences in the way that children and adults describe motion events correspond to differences in the way that they conceptualize those events. As expected, preschoolers in our study produced more limited descriptions of motion events than adults did, often

omitting details about the Path of motion depicted in our events. Beyond these linguistic differences, however, our data do not provide any evidence that would lead us to conclude that preschoolers' conceptual systems were any less sophisticated than adults.

On the contrary, our results demonstrate that despite differences in linguistic output, initial stages of utterance preparation are strikingly similar in preschoolers and adults. First, when preschoolers did include Path information, they did it in the same way that adult speakers of their language did—by adding a satellite specifying the Path to a verb specifying the Manner. In addition, eye movement patterns revealed that preschoolers and adults were prioritizing attention allocation to event components in highly similar ways. Specifically, both groups allocated more attention to event components that they planned to talk about. Patterns of eye movements in the Nonlinguistic task revealed that these similarities in event apprehension were independent of speech planning. And finally, even when they did not talk about them, preschoolers showed higher accuracy at identifying changes to Path endpoints in our motion events as opposed to Manners. Taking all of these results together, we conclude that preschoolers' tendency to omit path information from their linguistic descriptions was due to performance limitations rather than to inadequacies in event encoding.

References

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.

Barr, Dale. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.

Bartsch, Karen, & Wellman, Henry M. (1995). *Children talk about the mind*. Oxford University Press, New York.

Bloom, Lois. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Bloom, Paul. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 21*, 491–504.

Bock, Kathryn, Irwin, David E., & Davidson, Douglas J. (2004). Putting first things first. In J.M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 249–278). New York: Psychology Press.

Brown, Roger. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Casasola, Marianella, Cohen, Leslie B., & Chiarello, Elizabeth. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development*, 74, 1–15.

Fisher, Cynthia. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31, 41–81.

Fisher, Cynthia. (2002). Structure limits on verb mapping: The role of abstract structure in 2.5-yearolds' interpretations of novel verbs. *Developmental Science*, 5, 55–64.

Gentner, Dedre. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), *Language development, vol. 2: Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

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

Gleitman, Lila, January, David, Nappa, Rebecca, & Trueswell, John. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544–569.

Griffin, Zenzi. M., & Bock, Kathryn. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279.

Hitch, Graham J., Halliday, M. Sebastian, Schaafstal, Alma M., & Heffernan, Thomas M. (1991). Speech, "inner speech," and the development of short-term memory: Effects of picture-labeling on recall. *Journal of Experimental Child Psychology*, 51, 220–234.

Huttenlocher, Janellen, Smiley, Patricia, & Charney, Rosalind. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, 90, 72–93.

Johnston, Judith R., & Slobin, Dan I. (1978). The development of locative expressions in English, Italian, Serbo-Croation and Turkish. *Child Language*, 6, 529–545.

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

Palmer, Sue. (2000). Working memory: A developmental study of phonological recoding. *Memory*, 8, 179–193.

Papafragou, Anna, & Selimis, Stathis. (2010). Event categorisation and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, 25: 224–260.

Papafragou, Anna, Hulbert, Justin, & Trueswell, John C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108, 155–184.

Perner, Josef, Sprung, Manuel, Zauner, Petra, & Haider, Hubert. (2003). 'Want that' is understood before 'say that,' 'think that,' and false belief: A test of de Villiers's Linguistic Determinism on German-speaking children. *Child Development*, 74, 179–188.

Piaget, Jean. (1955). *The child's construction of reality*. London: Routledge and Kegan Paul.

Pinker, Steven. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Poulin-Dubois, Diane, & Graham, Susan A. (2007). Cognitive processes in early word learning. In E. Hoff & M. Shatz. (Eds.), *Blackwell handbook of language development* (pp 191–211). Malden, MA: Blackwell Publishing.

Quinn, Paul C. (2005). Developmental constraints on the representation of spatial relation information: Evidence from preverbal infants. In L. Carlson & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization, and development* (pp. 293–309). New York: Oxford University Press.

Rice, Mabel L., Redmond, Sean M., & Hoffman, Lesa. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research*, 49, 793–808.

Slobin, Dan I. (1973). Cognitive prerequisites for the development of grammar. In C.A. Ferguson & D. I Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York: Hold, Rinehart & Winston.

Slobin, Dan 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.

Talmy, Leonard (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.

van den Weijer, Joost. (1999). *Language input for word discovery*. Doctoral dissertation, Nijmegen: Katholieke Universiteit (MPI Series in Psycholinguistics, 9, Max-Planck-Institute for Psycholinguistics, Nijmegen).