

New Perspective for Verb Learning

Hanako Yoshida (yoshida@uh.edu)

Department of Psychology, University of Houston, 126 Heyne Building
Houston, TX 77204-5022 USA

Linda B. Smith (smith4@indiana.edu)

Psychological and Brain Sciences and Program in Cognitive Science, Indiana University
Bloomington, IN 47405 USA

Brian Weisinger (brian.weisinger@hotmail.com)

Department of Psychology, University of Houston, 126 Heyne Building
Houston, TX 77204-5022 USA

Abstract

When technology eventually allows us to zoom in to the point of view of children, will it provide us with any informative information about what the early word-learning environment really looks like? The present study uses a child-centered view that simulates infants' in-the-moment view to test whether such a newly obtained perspective may provide us with new insight into early verb learning. We present adult participants with two perspectives taken from child-parent play sessions in which parents were teaching the child a set of nouns and verbs. One was the child-centered view, and the other was a typical third-person view, and both were used without audio. The participants' task was to identify what words the parents taught by looking at the video clips. The results support higher accuracy in identifying nouns than verbs, and they provide a new finding about adults identifying verbs better than nouns when viewing the child-centered view.

Keywords: verb learning; word learning context; child-centered view.

Introduction

One of the most crucial questions in language learning is how children map words to meanings. Needless to say, there are a number of factors contributing to the process of word learning, including social cues such as eye gaze and gesture, prosody, language structure, input frequency, pragmatics, and much more. However, a complete explanation may require more than a characterization of the learning environment because learners come with biases. Moreover, learners actively engage in their world, and in so doing they distort the regularities and carve the input in systematic ways. This means that one cannot really consider the input separately from the learners' own actions because the learner selects and creates the input. Although the study of human cognitive development has focused on the "input" as divorced from the learner's activity, the growing interest in embodiment suggests the need to take a closer look.

One recent study that motivated this present study investigated how early learners create their own visual input by studying the first-person view during toy play (Yoshida

& Smith, in press). This study used a small camera attached to small children's foreheads and thus documented how this view may differ from what can be captured from a third-person view (commonly used for standard observation studies). In the present paper, we show how this child-centered view might change our understanding of early word learning, and particularly the relevant information for learning nouns and verbs.

Perceptual Accessibility and Word Learning

Gentner (1978) pointed out that there are perceptual differences in the accessibility of meanings among different word categories that determine how easily the words can be learned. In particular, she suggests that nouns dominate verbs in early productive vocabulary (see also Gentner 1982; Gentner & Boroditsky, 2001) because noun meanings are more available from the perceptual input. The natural partitions part of their hypothesis further proposes that animates and objects present cohesive perceptual properties easily separable from complex scenes. These cohesive perceptual bundles are most likely lexicalized as nouns, and as a result, they argue, young word learners universally learn nouns easily (see also Gillette et al., 1999). Verbs, in contrast, being about relations, are not easily specified by the perceptual input.

Human Simulation Paradigm (HSP)

The effect of perceptual accessibility was made dramatically apparent in a study of adults' ability to interpret the referents of real child-directed utterances of parents (Snedeker & Gleitman, 2000). The evidence comes from what is known as the Human Simulation Paradigm (HSP). In these studies, adults are asked to guess a set of words given visual information of child-mother play. The evidence shows that their rate of successful guessing differs depending upon the word type (Snedeker & Gleitman, 2000; see Piccin & Waxman, 2007, for a similar result with children; see also Snedeker, Gleitman, & Brent, 1999; Gillette, Gleitman, Gleitman, and Lederer, 1999; and Snedeker, 2000). In brief, they are better at guessing nouns from such scenes than verbs.

The key findings are from experiments in which adult participants watch a video of a scene of mother and child at play and are asked to identify a target word in the parent's speech. These are words the adults already know, of course, but they are disguised as nonsense words, beeps, or tones, and often other information about the linguistic context is removed (see also Snedeker et al., 1999; Gillette, et al., 1999; Snedeker, 2000). The results were highly consistent: adults were better able to identify nouns, but the context did not provide adequate information for adults to identify verbs (Gillette et al., 1999; Snedeker et al., 1999). This finding has been considered to be clear-cut support for the obviousness of perceptual information relevant to determining an intended noun referent on the one hand, and the non-perceptual and less accessible nature of verb meanings on the other. Although the conclusion that nouns are easier to learn than verbs is not in doubt, and it is likely the case that noun categories are more strongly constrained perceptually, the results from the human simulation experiments may be fundamentally misleading about the information available to and used by human learners in learning verbs.



Figure 1: Typical third-person view used for HSP.

Child Centered View

The HSP experiments record a view of a parent interacting with a child as shown in Figure 1. This is the third-person view of the interaction, the view of the observer, not the view of the child. Could it be that information about nouns and objects depend less on the perspective of the observer than on relational information about verbs?

To answer this question, we placed a head camera on children as they interacted with their parents, recording their first-person view of the scene (Figure 2, right). We also recorded the interactions from the usual third- person view (Figure 2, left.) We then replicated the HSP procedure, comparing adults' ability to determine the target noun or verb from the first- versus third- person views.



Figure 2: Snapshot of stimuli movie clips showing the first-person view (right) and the third- person view (left).

More specifically the stimuli were movie clips taken from a prior head camera study of mother-child interactions (Yoshida & Smith, in press). In that work, the system was specifically designed to record the visual input from the child's point of view. This was done by asking the child to wear a light-weight wide angle camera (90 degrees) worn low on the forehead. The camera thus recorded the event from the child's perspective. The context in which Yoshida and Smith collected their coupled third- and first-person views was that of naturalistic word learning in the context of toy play. The parent was instructed to teach the child a set of four nouns (cookie, bunny, car, bottle) and four verbs (open, eat, drink, put), by selecting appropriate toys from a large set of available toys. The dynamic first-person perspective is substantially different from the third-person (experimenter) view commonly used in developmental studies of the learning environment in that the structure and content of the first-person view is intimately related to the child's own goal-directed action. The present study selected videos from four participants in the original study in such a way that each dyad presented good first- and third-person views of all 8 words. These videos were then presented to adults, who were asked—in various ways—the target word.

Experiment 1

Here, we use an open procedure in which participants are not told the syntactic category of the target word. In this way, we first address the question of whether adults are biased to assume that the target word—what the mother is likely to be talking about—is a noun, regardless of the actual target, and whether any such bias is stronger for first- or third-person views. Participants were shown each clip and simply asked to provide three words for each clip that were good candidates for the target word. Does the child-centered view yield more accurate judgments? Does it matter more for verbs than nouns?

Method

Participants Fifty-four college students from the University of Houston participated in the study. The participants were randomly assigned to the two conditions: (1) the first-person view (child-centered view), and (2) the third-person view.

Stimulus Materials The stimuli movie clips each lasted about 1 minute and were taken from four different parent-

child pairs, each teaching four nouns and four verbs. Figure 3 provides simultaneous frames from the first- and third-person views for the verb “eat.” For each condition, whether first person or third person, a set of movie clips was constructed consisting of 24 movie clips—3 dyads X 8 typical early learned words—4 nouns (open, bunny, car, bottle), and 4 verbs (open, eat, drink, put). The order of the movie clips was randomized for each participant in each condition, and they were played on a 21-inch computer screen.

Procedure The participants were told to first watch each movie clip and provide three words that could be the word the mother is teaching by filling in the three slots on the response sheets. The participants were instructed to repeat the procedure for all 24 movie clips, at a comfortable speed. There was no audio, thus participants’ responses were solely based on the visual component if the events. The only difference between the conditions was the perspective of the view of the teaching event—the child’s view or an outside observer’s view.

Results

All reported effects were significant at the $p < 0.05$ level, and unreported effects were not significant. The first result, which is evident in Figure 3, is that participants are more accurate given the first-person view. A 2 (view type: first person, third-person) by 2 (word type: nouns, verbs), the word type as within subjects’ factor, analysis of variance revealed a significant main effect of view type: adult participants identified target words more accurately when viewing a first-person view ($M = 83.5\%$) than when viewing a third-person view ($M = 68.5\%$).

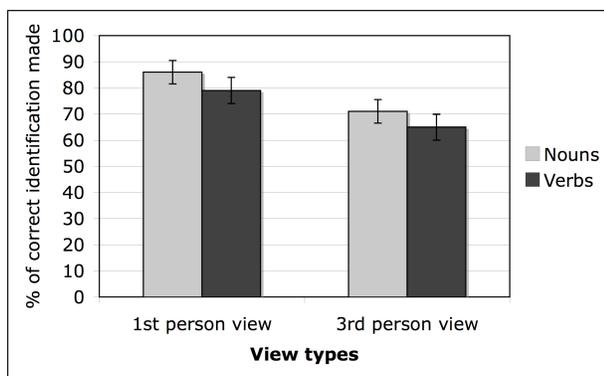


Figure 3: Word identification performances (% of correct identifications made) from Experiment 1.

In contrast to the original HSP study, no noun advantage was found for either view. In this free word-listing task context, adult participants did list more nouns than verbs, and in this sense they found each scene more compatible with more nouns (a noun disadvantage?) than with verbs. These observed differences from previous studies may well derive from this method of generating several possible

words unrestricted by syntactic category. Experiment 2 addresses this question.

Experiment 2

In this experiment, we aimed to test the effect in a more constrained context like those used in previous HSP studies. Here, we controlled the number of nouns and verbs to be identified. Participants were told that the movie clip showed either a noun-learning context or a verb-learning context, and they were instructed to generate only 1 word of a specified word type.

Participants. Seventy-nine college students from the University of Houston participated in the study. The participants were randomly assigned to the two conditions: (1) the first-person view (child-centered view), and (2) the third-person view.

Stimulus Materials and Procedure The stimuli and procedure are exactly the same as those used in Experiment 1, except for the way in which participants were instructed to identify the target words. In Experiment 2, adult participants were provided with a correct word type for each movie clip for them to identify one word. For example, take the scene in which a mother is teaching the child the word “open”; for that trial, participants were told to identify a verb.

Results

All reported effects were significant at the $p < 0.05$ level, and unreported effects were not significant. There is a reliable verb advantage in the first-person viewing condition. A 2 (view type: first person, third-person) by 2 (word type: nouns, verbs), the word type as within subjects factor, analysis of variance revealed two significant main effects as well as the interaction. Overall, as can be seen in Figure 4, adult participants identified target words more accurately when viewing a first-person view ($M = 75.2\%$) than when viewing a third-person view ($M = 61\%$). The reliable interaction suggests that the effect of view type is limited to the first-person view.

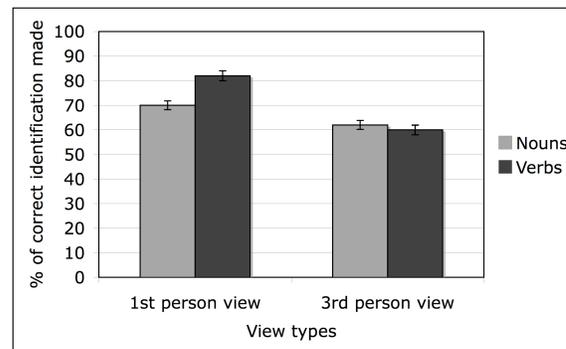


Figure 4: Word identification performances (% of correct identifications made) from Experiment 2.

The results suggest that a first-person view—the view of the child learner—aids in identifying the verb. This in turn suggests that the perspective of the learner may be particularly critical to verb learning. While objects are objects no matter where or from what direction they are viewed, verbs are about relations between agents and objects, and here point of view matters.

General Discussion

The most important contribution of these results is that they remind us that the relevant input for learning is only the input that the learner actually receives. Visual input depends upon the orientation of the head and eyes; the body's momentary disposition in space selects and filters the information and gives it a direction with respect to the viewer. Conclusions about what can and cannot be learned from the available input in any domain cannot be made without taking into account the view of the learner. Direct access to the first-person view—through the use of head cameras, for example—should greatly enrich our understanding of the information upon which children's learning is based.

In the context of noun and verb learning, we do not conclude that there is perceptual information available to learners, which is just as constrained and universal for verbs as for nouns. This seems unlikely on many grounds, but we do conclude that serious debate about the nature of the information available to early noun and verb learning requires a more serious examination of just what the perceptual information is, and, in the case of verbs, the perspective of the learner and how that perspective highlights some information and excludes other information may be deeply informative. There is much to be done in analyzing the information in these first-person views with respect to the meanings of the few verbs examined here and also, in future studies, across a wider range of verbs and kinds of meanings. The present results, however, strongly suggest that this is a task worth undertaking in that there is information in the first-person perspective that helps adults determine the target word that is not there in the third person view—information more critical to identifying a verb than a noun.

Acknowledgments

The video stimuli from the previous study made this work possible, and we thank NIH grant R21EY017843-01 for supporting the study. We thank the University of Houston for their support for data collection and coding for the present study.

References

- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S.A. Kuczaj, II (Ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 31–53). Hillsdale, NJ: Erlbaum.
- Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S.

- Levinson (Eds.), *Language acquisition and conceptual development*. Cambridge, U.K.: Cambridge University Press.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulation of vocabulary learning. *Cognition*, 73, 135–176.
- Gleitman, L. (1990). The structural sources of word meaning. *Language Acquisition*, 1, 3–55.
- Piccin, T. B., & Waxman, S. R. (2007) Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, 3(4), 295–323.
- Snedeker, J. (2000). Cross-situational observation and the semantic bootstrapping hypothesis. In E. Clark (Ed.), *Proceedings of the thirtieth annual child language research forum*. Stanford, CA: Center for the Study of Language and Information.
- Snedeker, J., & Gleitman, L. (2000). Why it is hard to label our concepts. Unpublished manuscript, University of Pennsylvania, Philadelphia, PA.
- Snedeker, J., Gleitman, L., & Brent, M. (1999). The successes and failures of word-to-world mapping. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the twenty-third Boston University conference on language development*. Somerville, MA: Cascadilla Press.
- Yoshida, H., & Smith, L. B. (in press). Using a head camera to study the first-person view of social interactions. *Infancy*.