On the Feasibility of Cross-situational Learning of Verbs

Researchers have long speculated on the role of cross-situational learning for language (Frank et al., 2009; Quine, 1960; Yu & Smith, 2007). However, both theoretical considerations and empirical results have raised questions about how useful cross-situational learning is for *verbs* (Gleitman, 1990; Medina et al., 2011; Trueswell et al., 2013).

Several studies show that verbs (especially manual and action verbs) often (not always) describe events that happen within half a minute or so of the utterance (Liu et al., 2019; Rodrigo et al., 2020; West et al., 2022). However, many other events presumably also happen during this window, making the evidence highly ambiguous (Gleitman, 1990; Quine, 1960). Prior work has not quantified just how ambiguous.

We identified uttered verbs (tokens=389; types=59) from two videos of mother-infant dyads from the Rollins corpus (*me06.cha* and *di06.cha*) in CHILDES (MacWhinney, 2000; Rollins, 2003). Consistent with prior findings, 34% of verbs referred to an event within a 10-second window around the verb, and 53% referred to one within a 50-second window (45% did not refer to any event within the video recording). See Fig. 1a.

What events the learner might be *considering*? Two groups of annotators labeled events in 10-second, 30-second, 1-minute, and 2-minute silent clips (different lengths highlight different sorts of events). The first group was asked to label X "important and salient" events in each clip. Despite identifying 8,408 unique events, the coders only noted the corresponding event for 13% of verbs (24% of those with identifiable events).

Perhaps infants rely on social cognition (Tomasello, 2010). The second set of coders labeled events "worth talking about". They identified 10,852 unique events, which included the corresponding event for 26% of verbs (48% of those with identifiable events), a substantial improvement. However, even so, the target events were a negligible fraction of events, a problem that increases rapidly the wider time interval considered (Fig. 1b).

Cross-situational learning is of limited help. Each verb type was uttered between 1 and 39 times (Mean=6). Considering events within a 10-second window of the verb, 53% of verbs *never* co-occurred with the target event type. For verbs that did, a median of 29 other event types were at least as associated with the verb (range=5-106). Increasing the window *decreases* the relative association between verb and target event type.

One issue is that many event types were very common (285 event types occurred in at least half of the 10-second intervals) and are thus highly associated with all verbs. Filtering these out does not help, though, since many of the verbs do in fact refer to these event types [Fig. 1c; event frequency was significantly if moderately correlated with verb frequency ($r=.15$, $p=.01$)].

We discuss limitations of the dataset and analyses. Nonetheless, results suggest a limited role in verb learning for verb-event co-occurrences. We discuss these findings in the context of "eureka" moments (Medina et al., 2011) and mechanisms for sharply filtering out possible verb meanings (Gleitman, 1990; Tomasello, 2010).

500 words

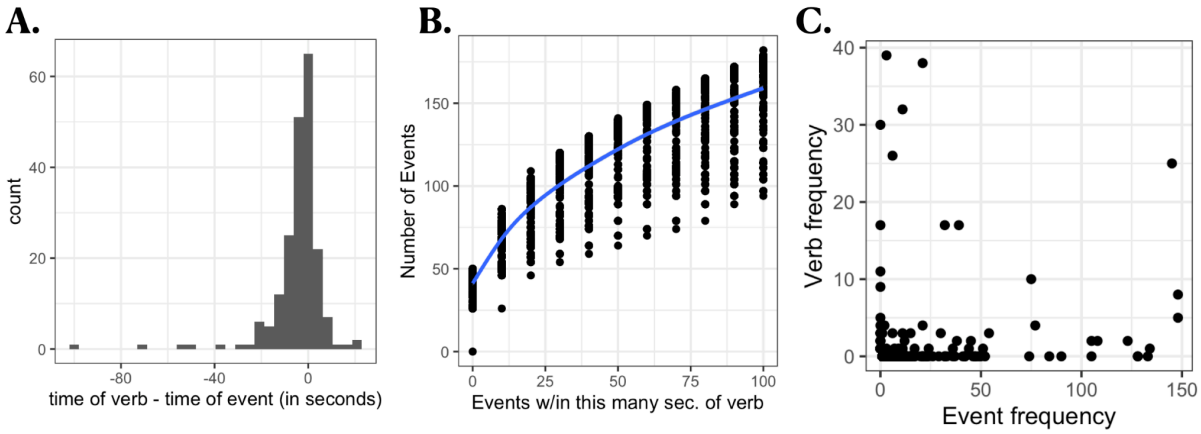

Figure 1. **A**. Histogram of temporal distance between verbs and the event they describe (if any). **B**. Number of distinct event types within an X-second window of verb, averaged over verbs. **C**. Frequency of event types (x-axis; maximum possible is 148) against frequency of verbs (y-axis)

References:

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological science*, 20(5), 578-585. **Gleitman, L.** (1990). The structural sources of verb meanings. Language acquisition, 1(1), 3-55. **Liu, S., Zhang, Y., & Yu, C.** (2019). Why Some Verbs are Harder to Learn than Others–A Micro-Level Analysis of Everyday Learning Contexts for Early Verb Learning. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 41). **MacWhinney, B.** (2000). *The CHILDES Project: Tools for analyzing talk: The database* (3rd ed.). Lawrence Erlbaum Associates Publishers. **Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R.** (2011). How words can and cannot be learned by observation. Proceedings of the National Academy of Sciences, 108(22), 9014-9019. **Rodrigo, M. J., Muñetón-Ayala, M., & de Vega, M.** (2020). Exploring the Co-occurrence of Manual Verbs and Actions in Early Mother-Child Communication. Frontiers in Psychology, 11, 596080. **Rollins, P.R.** (2003). Caregiver contingent comments and subsequent vocabulary comprehension. *Applied Psycholinguistics, 24*, 221-234. **Tomasello, M.** (2010). Origins of human communication. MIT press. **Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R.** (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology, 66*, 126- 156. **Quine, W. V.** (1960). *Word and object*. MIT Press. **West, K.L., Fletcher, K.K, Adolph, K.E., & Tamis-LeMonda, C.S.**(2022). Mothers talk about infants' actions: How verbs correspond to infants' real-time behavior. *Developmental Psychology, 58*(3), 405-416. **Yu, C., & Smith, L. B.** (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, 18(5), 414-420.