In E. Clark (ed.), *Proceedings of the Thirtieth Annual Child Language Research Forum.* Stanford, CA: Center for the Study of Language and Information (2000).

Cross-Situational Observation and the Semantic Bootstrapping Hypothesis*

JESSE SNEDEKER

1 Word learning as word-to-world mapping

How do children learn their first words? Everyone who has thought about the problem long enough to consider one has come up with essentially the same solution. Initially learners must: pair each word with the extralinguistic contexts in which it occurs, collect several such pairs, and then identify the common element in the scenes. We may disagree on the parents' role in structuring the input or the range of hypotheses that the child entertains but we all agree on the data source. After all, what other information does the novice language learner have?

But reasonable people do disagree about how much children can or must learn in this manner. Both empiricist philosophers (Locke, 1690; Hume, 1758) and nativist linguists (Wexler & Cullicover, 1980; Pinker, 1984) have suggested that children use extralinguistic contexts to learn the meanings of a large and diverse set of words. In contrast, the proponents of syntactic bootstrapping claim that this word-to-world mapping procedure can only provide the child with a small starter lexicon, consisting mostly of concrete nouns. This initial vocabulary is a wedge that the child uses to begin constructing representations of the sentences in which novel words are used. These representations constrain her interpretation of extralinguistic context, allowing her to acquire verbs, adjectives and abstract nouns

* The analyses presented here benefited greatly from conversations with Saul Sternberg, Michael Kelly, Michael Brent, Henry Gleitman, Lila Gleitman, and the members of the Cheese Seminar. This work was supported by an NSF Center Grant to the University of Pennsylvania Institute for Research in Cognitive Science

(Gillette, Gleitman, Gleitman, & Lederer, in press; Snedeker, Gleitman, & Brent, 1999).

To settle this issue we will ultimately need to determine: how much evidence about word meaning is available in word-to–world pairings; whether young children can extract and represent this evidence; and to what extent they actually it use it. The work described here only attempts to address the first of these questions.

2 The Human Simulation Paradigm

This first question is one about the correlation between two types of information: the properties of the scene and aspects of word meaning. Computer simulations are often used to study questions of this type. Essentially any machine capable of representing both types of information can be used to analyze their correlation. In the studies I will be describing, we used adult "learners" to conduct this analysis. Adults can represent both scenes and word meanings, and their representations are apt to be more like children's than any limited set of coding categories we might create.

In the Human Simulation Paradigm, we ask adults to identify a word, based on information about the contexts in which it was used in infant-directed speech. Thus, while the learners are artificial, the input that they get is authentic. By filtering the context in various ways, we can explore the relative potency of different information sources or the ease with which different types of words can be learned (Snedeker & Gleitman, in press). Previous studies have found that, while extralinguistic contexts provide adequate information for the identification of nouns, they do not provide enough information to allow adults to identify verbs (Gillette et al., in press; Snedeker et al., 1999). This finding has been offered up both as an argument supporting the need for structural information in verb learning and as a possible cause for the object-category bias in early word learning.

The analyses and arguments that follow are based on the data collected in Snedeker et al. (1999). In this study, the target words were the twenty-four most common nouns and twenty-four most common verbs from videotaped transcripts of four hour-long conversations between mothers and their children (18-24 months). Six instances of each word were selected at random and a videoclip was constructed that began thirty seconds before the mother used the word and ended ten seconds after she did so. The audio was removed from the videoclip and a tone was inserted exactly where the target word had been used (for details see Gillette et al., in press).

The words were divided into three lists and twenty-eight subjects were shown the videoclips for each list. The subjects were fully informed about their task (to identify the word the mother had said) and the source of the materials (they knew that the words were the most common nouns and

verbs from maternal speech). However, they were not told the syntactic category of the target. They attempted to identify the target word after each tone and then, after viewing all six clips, they looked over their previous responses and made a final conjecture. On the final trial, they succeeded in identifying 26 percent of the noun targets but only 12 percent of the verbs.

3 The Semantic Bootstrapping Hypothesis

This finding is problematic for language acquisition theories that rely heavily on the efficacy of a word-to-world mapping procedure. Here I will focus on one theory with this property, the semantic bootstrapping hypothesis (SBH).[1] SBH is a proposal for how semantic categories are used to bootstrap syntactic categories so that innate syntactic knowledge can be applied (Pinker, 1984). According to this proposal, the learner: 1) acquires a sizeable and diverse vocabulary on the basis of cross-situational observation; and then 2) uses universal syntactic-semantic correspondences to determine the syntactic category of each word. If verb meanings cannot be learned from extralinguistic contexts, then the learner will be unable to identify instantiations of this category in the input and therefore will be unable to apply any syntactic knowledge relating to verbs.

But before we conclude that verbs cannot be learned from extralinguistic contexts, it is critical to point out two questionable properties of the Human Simulations. First, the adult learners were given only six contexts for each word, while real children receive thousands of exposures. Perhaps adults failed to identify verbs, not because they cannot be learned from this information source, but because they were not given enough examples to allow them to do so. Second, previous analyses of the data have focused on whether subjects could correctly identify the target words, which presumably requires the adults to learn enough about each target to distinguish it from all other candidates. Semantic bootstrapping does not demand such perfect knowledge of word meaning. This paper addresses these concerns by conducting new analyses of the data from Snedeker, Gleitman, & Brent (1999). Section 4 examines the learning curves and patterns of convergence to evaluate whether the noun-verb difference is likely to hold up as the number of learning trials increases. Section 5 looks more closely at the incorrect responses. When the extralinguistic context fails to allow subjects to identify the word, does it provide partial information that could guide semantic bootstrapping?

---

[1] Extralinguistic context is also the primary information source for word learning in the theories of many modern day empiricists and social interactionists. The problems that the Human Simulations raise for these proposals are laid out in Snedeker & Gleitman (in press).

4 Could additional contexts improve verb performance?

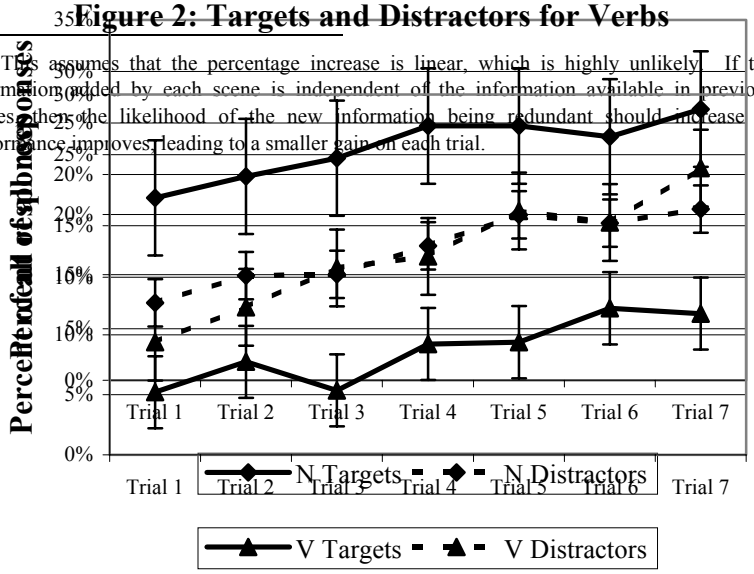
The subjects in these studies were given six contexts for each word. This number was arbitrary, a compromise between ecological validity and subject fatigue. The real child is exposed to each of these common words thousands of times before she uses them. Thus it is possible that the child gradually learns the word from an accumulation of evidence and only arrives at a precise identification of its meaning after dozens or even hundreds of exposures. Perhaps, despite the noun-verb difference, word-scene pairs contain enough information to identify all words, but only if a very large number of pairs are provided.

Further analysis of the changes in performance across trials provides some support for this conjecture. As suggested in Figures 1 and 2, the learning curve is steeper for noun targets than for verb targets, resulting in a significant interaction between Syntactic Category and Trial ($F = 2.7$, $p < .001$). However, for both target types there was a significant effect of Trial, indicating that performance improved as the subjects were given more word-scene pairs ($F = 9.6$, $p < .001$ for nouns and $F = 17.3$, $p < .001$ for verbs). If each additional scene lead to similar improvement, then performance on both nouns and verbs would be perfect by the eighty-fifth trial.[2]

But another pattern in the data suggests that this forecast is highly implausible. While subjects are more likely to identify the correct verb as more information accumulates, they also tend to converge on the same incorrect responses. Obviously, both of these processes cannot continue indefinitely. The ability to identify a word requires both that the input contains cues that point to the correct meaning and that these cues are more systematic or salient than the cues pointing toward other candidates. If the potency of misleading cues increases faster than that of accurate cues, then additional contexts will not improve performance. We compared the strength of these two competing tendencies by: 1) identifying the most common incorrect answer for each target on the final trial (the "Distractor"), 2) calculating the number of subjects who gave that response for each trial of that target; 3) comparing the Distractor and Target totals in an items ANOVA. There was a strong interaction between Syntactic Category and the Distractor-Target variable

Figure 1: Targets and Distractors for Nouns

Figure 2: Targets and Distractors for Verbs



[2] This assumes that the percentage increase is linear, which is highly unlikely. If the information added by each scene is independent of the information available in previous scenes, then the likelihood of the new information being redundant should increase as performance improves, leading to a smaller gain on each trial.

($F = 7.0$, $p < .05$). For nouns, Targets were more common than Distractors from the first trial to the last (Figure 1). For verbs, however, this pattern was reversed (Figure 2). Distractors began with an advantage that only grew larger between the first and final trial. Word-scene pairs not only provided less evidence about verb meanings than about nouns, they actually contained systematic misinformation, red herrings which lead our subjects to converge on false hypotheses.

5 Could partial knowledge of verb meaning guide semantic bootstrapping?

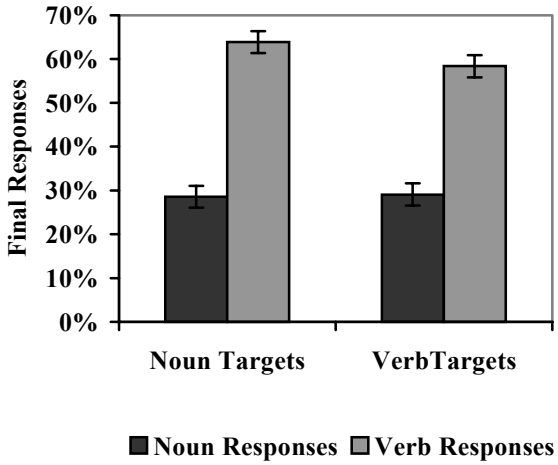
The SBH does not require scenes to provide complete and accurate semantic representations. The child only needs to learn the syntactically relevant aspects of word meaning in order to begin identifying syntactic categories. In previous analyses of these data, responses were only counted as correct only if the subject identified the root morpheme of the target. For example, if the target word was *look* we accepted *looked* or *looking* but not *see*. Thus a subject could learn a great deal about a word but still not be credited with identifying it. The following analyses look at whether the incorrect responses overlapped in meaning with the targets more than would be predicted by chance alone.

At the coarsest level, subjects might be able to determine whether a word was a noun or a verb even when they could not identify the word itself. To test this hypothesis, the correct responses were removed from the data and the remaining responses were categorized as nouns, verbs, or others.[3] The number of noun responses was summed separately for noun and verb targets and submitted to both a subjects and an items ANOVA. As shown in Figure 3, there is no effect of the syntactic category of the target on the syntactic category of the response ($F1 < 1$, $p > .8$; $F2 < 1$, $p > .4$). When subjects failed to identify a word, they apparently had no information about whether it was a noun or a verb.

The next analysis focused on the verb targets. The SBH proposes that children use semantic information, gleaned from the real world context, and innate knowledge of universal syntactic-semantic correspondences to identify how different structures are instantiated in the language they are learning. One of the primary motors of development according to this story, is the child's ability to use a verb's meaning to predict the types of structures that it can occur in. To do this, the child need not learn everything about a

[3] At the end of the session the experimenter identified words that were potentially ambiguous and asked the subject if each one was "an action or an object".

Figure 3: Syntactic Category of Incorrect Responses for Noun and Verb Targets



given verb. She only has to discover enough to place it in the correct syntactic-semantic class. In this analysis, we examined whether the incorrect responses of our adult subjects suggested that they had any information about the verb class of the unknown targets.

To do this we first assigned each target verb and incorrect response to one of Levin's (1993) verb classes.[4] Next, for each verb we calculated: 1) the percentage of incorrect responses that came from the same class as the target (hits); and 2) how often verbs from this class appeared as incorrect responses for all unrelated targets (false alarms). These two percentages were entered into an items ANOVA. Although the mean percentage of hits was somewhat higher than that of false alarms (4.7 vs. 3.6), this difference did not approach significance ($F < 1$, $p > .5$) and appeared to be due to above chance performance on just two targets (*come* and *see*).

[4] In this analysis the broadest classes from Levin (1993) were used. However, an analysis using the finer subdivisions produced similar results. One verb (*do*) could not be classified and was therefore excluded from the analyses.

6 Summary

More fine-grained analyses of the Human Simulation data confirm that extralinguistic contexts contain adequate information for learning nouns but not verbs. Noun scene sets strongly support word learning and generate few false hypotheses. Verb scene sets do not provide enough information to allow subjects to identify the target word and lead them to converge upon false hypotheses. Close examination of the incorrect responses indicates that when subjects fail to identify the word, they apparently learn little about it; they cannot determine the syntactic category of the target and, if the target is a verb, they do not learn which verb class it belongs to. These results strongly suggest that a word-to-world mapping procedure could not produce the syntactically diverse lexicon that the SBH requires. However, this noun-advantage is consistent both with the known facts about early vocabulary composition and with a model of word learning in which word-to-world mapping is rapidly supplanted by sentence-to-world mapping as the child gains a linguistic foothold.

References

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (in press). Human Simulation of Vocabulary Learning. To appear in *Cognition*.

Hume, D. 1758. *An Inquiry Concerning Human Understanding*. Indianapolis, Ind.: Bobbs Merrill, 1955.

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

Locke, J. 1690. *An Essay Concerning Human Understanding.* 259-98. Cleveland: Meridian Books, 1964.

Pinker, S. 1984. *Language Learnability and Language Development*, Cambridge, MA, USA: Harvard University Press.

Snedeker, J. & Gleitman, L. (in press). Why it is so Hard to Label our Concepts. To appear in Carey, S., Bonnatti, L., & Mehler, J. (eds).

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.

Wexler, K. & Cullicover, P. 1980. *Formal Principles of Language Acquisition*. Cambridge: MIT Press.