Whorfian Biases in the Categorization of Events

Ángel Cabrera and Dorrit Billman
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170
e-mail: angel@psy.gatech.edu

Abstract

The purpose of this paper can be viewed from two different but related perspectives. First, it investigates the role of language in concept learning. Second it investigates this by asking how event categories are acquired. From the first perspective, our experiment showed that syntactic Whorfian biases in human categorization actually exist. The new methodology employed provides us with a tool to study in more detail the nature of those biases. The Whorfian hypothesis does not necessarily have to be tested cross-culturally. It can successfully be taken into the lab with subjects belonging to the same linguistic community. From the second perspective, our experiment showed that the addition of verbal descriptors to a set of animation scenes in some systematic way facilitates learning of the regularities in the scenes. This result is a new piece of evidence supporting the focused sampling theory of category acquisition.

Introduction

How are event categories learned and what is the role of language in such learning? Prior work on observational learning of "verb-sized" events assessed how varying the structure in the scene affected learning of the event categories. The current study extends this by varying the structure of the language describing the scene and asking how this affects learning the event categories. This is one test of the Whorfian Hypothesis (Whorf, 1956; Hunt & Agnoli, 1991), in which we ask whether and how language affects categorization of nonlinguistic events. We consider how questions about form and meaning have surfaced in recent work on language acquisition, we review work on learning event (or verb) categories, and summarize our approach to category learning. We then describe how the present study explores effects of language, both words and syntax, on learning event categories.

Whorf proposed that the syntax and lexicon of a language have a profound effect on thought. The semantic distinctions obligatory in syntax or marked by words affect the distinctions prevalent in the thinking of the language's speakers. Recently, problems in language acquisition have again raised the question of how form and meaning, language and thought, are related. Advocates of "semantic bootstrapping" propose that language learners use semantic correlates of syntax to aid syntax acquisition (Grimshaw, 1981). Advocates of "syntactic bootstrapping" propose that language learners use syntactic correlates of verb meanings to aid verb learning (Gleitman, 1990). Specifically, the syntactic bootstrapping hypothesis claims knowledge of syntax might aid the learning of the mappings of verbs onto the underlying concepts they denote. Both proposals address how learning one aspect of language can aid another aspect, not how language affects nonlinguistic aspects of cognition. But like Whorf, both approaches ask how the form of language is related to semantic or conceptual structure. Our work asks how knowledge of syntax might help learn the underlying event categories to which verbs refer. In addition, we investigate effects of lexical information and conditions with no language at all. We ask how information from the argument structure of English affects learning novel event categories.

Few studies have investigated learning of event categories, in contrast to the extensive body of work on object categories. Further, it is verb structure, not event categories *per se*, which is often the focus of analysis. While few comparisons of learning event versus object categories have been done, several researchers have suggested important differences, not just in the content of categories, but in the organizing principles as well. In particular, "good" object categories have rich correlational structure, while event categories may have a matrix-like organization. Many object categories have multiple, correlated properties that members of the category share but distinguish them from contrasting categories (Rosch, 1978). Verbs, and the event categories they designate, may be different in

---

This work was supported by a Fulbright/MEC Scholarship from the Spanish Ministry of Education and Science and the Institute of International Education awarded to the first author.

several regards. Gentner (1981) pointed out that verbs differ from nouns in the way they are acquired, memorized and translated, in their bredth of meaning, mutability under paraphrase and in the cross-linguistic variability they show. Her explanation of these differences was based on the concept of *density* of internal links in the context of a semantic network representation of meanings. Noun meanings are more internally dense than verbs; nouns convey meaning by themselves. Verbs relate meanings. The meaning of a verb generally makes sense only in the context of the meanings it relates. Huttenlocher and Lui (1979) found higher performance for nouns at all ages in their developmental experiments on memory for semantically related verbs and nouns. In their interpretation of these results, they argued for a matrix-like organization for verbs. Thus, we need to ask, not assume, how learning of event categories is like learning of object categories. Despite the suggested differences, we propose that the learning mechanism for both object and event categories will try to capture correlations or inferential structure in the input. For both, people should be more successful when the relationships to be learned are coherently organized.

Our particular hypothesis about what constitutes coherent organization comes from the focused sampling method in the Internal Feedback Model (Billman & Heit, 1988). This model was designed for unsupervised learning (no externally provided category labels) where learners must generate their own feedback internally. The core idea is that learners seek to recover the predictive structure in input by learning predictive patterns or correlational rules among attributes. As such rules or patterns are discovered, learners increasingly focus attention on the attributes that are proving predictive. As a result, an individual correlation that is part of a larger system of related attributes will be learned more easily than when that correlation occurs by itself. This predicted facilitation was tested and found for syntactic categories of artificial grammars (Billman, 1989) and for object categories (Knudson & Billman, 1990). Given the suggested differences between verb and noun meanings, it was not clear whether the same pattern would be found for event categories. Kersten and Billman (1992) tested for the predicted facilitation of additional correlated cues in learning categories of simple transitive events shown as computer animations. When attributes in a target correlation (between A1 and A2) also correlated with additional attributes (A3 and A4), the target correlational rule was learned more easily. Presence of additional correlated attributes facilitated learning a target correlation, contrary to the inhibition predicted by competitive learning models developed in the context of classical conditioning (Rescorla & Wagner, 1972). For unsupervised tasks, richer correlational structure facilitates learning component relations. This is true for the domain of events as well as objects.

The present work explores the effects of language on learning event categories. Can information provided in language function as an additional, facilitative attribute? We investigated simple forms of lexical and syntactic information. Suppose that two or more aspects of the events (such as agent manner of motion and change of state of a patient) covary, providing a basis of distinguishing event categories. If the lexical item, here a verb, covaries with these attributes, the choice of verb may serve as an additional relevant attribute and facilitate learning the relation between the two nonlinguistic attributes, as in Kersten and Billman (1992). Now consider the contribution that syntax might provide. We believe speakers within a linguistic community share biases about the aspects of interaction that are relevant to event categories and that these biases are based (in part) on the syntactic structure of the language. If subjects learning event categories are biased toward certain aspects of interaction by the syntax of accompanying sentences, a syntactic distinction may function as an extra relevant attribute. Moreover, systematicity between syntax and other relevant attributes in the training scenes should facilitate learning of nonlinguistic relations in the scenes.

The plausibility of this view depends on the presence of regularities linking verb meaning and syntax. We focus on the role of verb argument structure. Verbs differ in the kinds of grammatical structures in which they occur. For example, in English we can say and easily understand

• *Chris put all the papers on his desk.*

while we probably would have a hard time with

• * *Chris joked all the papers on his desk.*

While *put* usually occurs in structures like *NP V NP PP* (Noun Phrase, Verb, Noun Phrase, Prepositional Phrase), *joke* occurs in structures like *NP V* (*John is joking*). These structures are usually referred to as the *subcategorization frames* associated with those verbs. Fisher, Gleitman and Gleitman (1991) have experimentally addressed the mapping between subcategorization frames of some verbs and their meanings. They found robust correspondence between judged syntactic and semantic similarity of verbs. They argued that "the syntax of verbs is a quite regular, although complex, projection from their semantics" (p. 331). This work on established English verbs suggested that verbs with similar meanings have similar syntactic frames. Such a pattern may also bias a learner toward a particular type of meaning in learning new verbs.

Our experiments select syntactic forms of English for a 'dialect' that uses all new nouns and verbs. In the No-Language Control condition, subjects view scenes with no language of any sort. In the Informative-Lexicon condition, scenes are accompanied by English-like sentences where novel verbs are used to covary with the event categories. The Informative-Syntax condition preserves one simple set of correlations between syntax and meaning found in English so that syntactic form as well as lexical item covaries with event category.

Method

Subjects

Forty-eight college students (16 per experimental condition) served as subjects, receiving extra-credit for volunteering. All were native English speakers.

Stimuli

The visual stimuli consisted of animated scenes programmed with the software MacroMind Director on a Macintosh II computer and displayed on a high-resolution 13" color Apple Monitor. Language was presented auditorily on the speaker built into the computer.

The task was divided into two consecutive phases: a learning phase and a test phase. During the learning phase all scenes showed some regularities which subjects were asked to discover. Each scene in this phase involved two animate objects: one squared or rectangular which we will refer to as a *mugglet* and one rounded which we call a *diggy* (see Figure 1). All three mugglets shared the same behavior patterns, and so did the three diggies. The particular exemplars appearing in each scene were chosen at random among the three alternatives with no systematic correlation with the rest of attributes described below.

At the beginning of each event, the diggy was located in the center of the screen, while the mugglet occupied the extreme of one of the eight principal radii from the center, also chosen at random. Each event could be described as a sequence of three phases (that we will consider as the attributes of the events) each of which could take of one of three different forms (that we will consider as the possible values of the attributes):

1. Initiation. Either the mugglet, or the diggy, or both at the same time shake vertically at their original location.

2. Approach. The mugglet approaches the diggy either smoothly, or with a transversal vibration (i.e. orthogonal to the direction of movement), or with a longitudinal vibration.

3. Final run. After the mugglet and the diggy contact, the diggy moves away according to one of the following patterns: (a) the mugglet stops in the center of the screen and the diggy runs off the screen, (b) the mugglet chases the diggy while it runs off the screen, or (c) the diggy starts jumping at random to different locations on the screen followed by the mugglet.

All three attributes were perfectly correlated with each other: each type of initiation phase always corresponded with a certain approach and a certain final run. There were therefore three event categories to be learned each of which could be characterized as a combination of one of the three possible initiation phases, one of the three approaches and one of the three final runs.

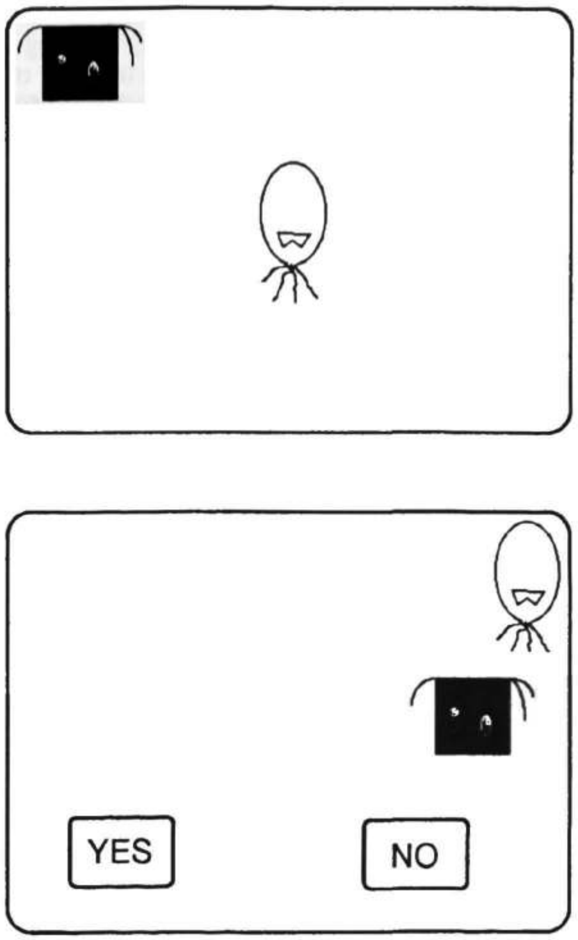

Figure 1. Initial and final state of a typical event. During the test phase, two buttons appeared on the screen at the end of the event.

In the conditions with language, a distinct verb was used for each of the three event categories. In the Informative-Syntax Condition, three different sentence types were used, one for each event category. The sentence types differed in syntax and mapping from syntactic argument to thematic role; either the agent is subject, the patient is subject, or there is a coordinate subject (see Table 1). We predicted that subjects would be biased toward having the subject of the sentence correspond with the object that moved first. This agency bias is reflected in the argument structures we chose for the Informative-Syntax condition.

Table 1. Scenes and Descriptive Sentences across Conditions.

Cat. #	Scene for All Conditions	Language in Informative-Syntax	Language in Informative-Lexicon
1	The mugglet vibrates goes to the diggy with transversal vibrations mugglet stops & the diggy flees:	*the mugglet troces the diggy*	*the mugglet troces the diggy*
2	The diggy vibrates the mugglet goes to it with a longitudinal vibration mugglet chases the diggy as it flees	*the diggy drokes the mugglet*	*the mugglet drokes the diggy.*
3	Both vibrate, the mugglet goes to the diggy smoothly diggy jumps; mugglet follows	*the diggy and the mugglet geel*	*the mugglet geels the diggy.*

In the test phase only the correlations between the visual attributes were tested; no spoken descriptors were provided. This phase consisted of a sequence of events similar to those in the learning phase, some of which showed the same correlations and some of which did not. There were 15 Consistent Event items, five for each of the three event categories. There were 15 Discrepant Event items that broke at least some of the rules. After each testing item, two buttons labeled as "YES" and "NO" appeared at the bottom of the screen. Subjects were asked to click either one indicating whether the event was consistent with those viewed during the learning phase.

Procedure

A pilot study with a reduced number of subjects per condition was run to find the number of learning trials likely to avoid floor and ceiling effects. Eighty subjects participated in either the No Language Condition or the Informative Syntax Condition at one of five possible number of training events (20, 30, 40, 50, 60, 70 or 80). The biggest difference was found at 40 training events which was finally chosen for the actual experiment.

After filling out the consent form, each subject was led to a computer in a sound-insulated booth and, when necessary, instructed in use of the mouse. Next, the subject was left alone and a set of instructions appeared on the screen. The subject had to click on the appropriate button to turn to the next page of instructions. The instructions told the subject that he was going to view images of events occurring in an imaginary world (PlutoZoo), described in the language of that world. He was also told that after viewing them, he would be asked to judge whether another sequence of events was a possible event in that world. Subjects were not asked to learn the language of that world.

Following the instructions, the first phase began. Presentation of scenes was self-paced. After the last learning event, there was a brief pause (5 seconds) and then the instructions of the test phase were presented in exactly the same way as the previous instructions.

The computer program recorded the number of correct answers for each type of testing item as *Inconsistent* (number of correct NO responses to inconsistent items) and *Consistent* (number of correct YES responses to consistent items from categories 1, 2 or 3) respectively.

Results

In general, *Consistent* was very high (in total, an average of 12.25 out of 15), therefore providing very little information. This suggests that subjects answered positively unless they detected some discrepancy. No significant difference across conditions was found for *Consistent.*

Fortunately, *Inconsistent* was a useful measure of learning performance. *Inconsistent* indicates the ability of the subject to detect events that disrupt the correlations present in learning and hence do not belong to any of the three original categories. The general pattern of distribution of the means for the three experimental groups are presented in Figure 2. Table 2 summarizes these numbers with their corresponding standard deviations.

Condition	Mean	SD
Inf-Syn	12.31	2.12
Inf-Lex	9.87	2.85
No-Lang	7.25	3.73

Table 2. ANOVA summary: "Inonsistent" on the three experimental groups

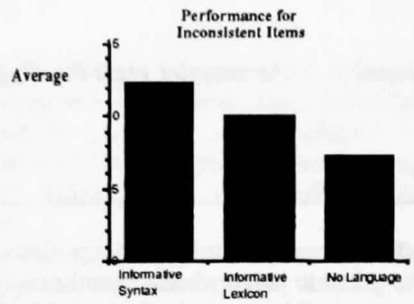

Figure 2. "Inconsistent" across the three experimental conditions in the final experiment.

The analysis of variance with *Inconsistent* produced a $F(2,45) = 11.591$, $p < 0.0001$ thus showing significant differences across the three experimental conditions. The two statistical comparisons planned beforehand also turned out to be significant for *Inconsistent*: subjects in the Informative-Lexicon condition performed better than subjects in the No-Language condition with a $t(45) = 2.496$, $p < 0.016$; and subjects in the Informative-Syntax condition performed better than the rest, with a $t(45) = 4.117$, $p < 0.001$.

The average performance of subjects in the group with no verbal labels (7.25, S.E.= .93) is not significantly different from the average score than one would get just by chance (7.5). The other two experimental conditions provide average *Inconsistent* scores that are significantly greater than the chance score.

After the task, subjects were asked to describe approximately the patterns of behavior they had discovered, and were encouraged to use the words used during the training phase (*mugglets, diggies, troce,* etc.) in their descriptions. Most subjects found it difficult to remember these words. All reported the existence of two types of characters (mugglets and diggies) with two different names but just a few could remember exactly what the names were. Verbs (geel, troce and droke) were even harder to remember. When the words were repeated by the experimenter, all subjects recognized them as being the labels used in the animation. Although it is not made clear during the experiment what *mugglet* and *diggy* refer to, all subjects in the two language conditions, assigned the nouns *mugglet* and *diggy* in the way we expected, i.e. in the way it was described above. Most subjects based their descriptions on analogies with social situations, interactions between opposite sexes being the most frequent metaphor.

Discussion

We hypothesized that language would influence learning about regularities in the scenes which form the basis for event categories. When lexical items, here the verb, covaried with event categories the relations determining the event categories were learned more quickly. This produced the better performance in the Informative-Lexicon condition over the No-Language condition. In the Informative-Syntax condition, syntactic form of the utterance itself --here the argument structure-- also covaried with event category. This syntactic information also served as a cue which subjects used to aid learning about the nonlinguistic world. In short, the Whorfian bias hypothesis was supported, in both the lexical and syntactic form.

While most studies on the influence of language on thought have been cross-linguistic, we propose an alternative pathway for exploring these phenomena. A new method has been presented to show that syntactic Whorfian biases actually exist. This method can be used to explore with more precision the nature of those biases.

Language played an active role in the learning of categories both lexically and syntactically. Lexical information helped to determine the number of different object categories and interaction patterns and to generate the internal feedback necessary for the learning process. Syntactic information directed the learner's attention toward a specific class of interaction patterns. Surprisingly, even subjects that had learned all the regularities in the scenes had difficulties retrieving the lexical items. They may have been used during learning, but forgotten since they were not needed for communication. Alternatively, since they were not rehearsed during the testing phase, remembering lexical items may be harder than recognizing the correlational rules that had been present during both phases. Testing subjects right after the learning phase would probably be the only way to address this.

Not all the theories about language and category acquisition would predict the benefits of language we found. Accounts of word learning often assume that concepts are acquired before the word, which is learned as a mapping onto an established category (see Huttenlocher, Smiley & Ratner, 1983). While it seems intuitive to think that providing a label should help learning somehow, the way this facilitation takes place is not so intuitive. For instance, it could be argued that verbal labels act as feedback in the learning task. The learner would be trying to find some criterion to predict each label. According to some competitive models of learning (Rescorla & Wagner, 1972), multiple cues would compete for predictive success of the label, and relations between the cues would not be learned. Our experiment showed that regularities between the attributes of the scenes were in fact learned. Moreover, presence of language facilitated this learning instead of impairing it. Syntax, too, was not only used to learn verb meanings (i.e. to predict the verb label), but to guide attention in the realm of finding regularities in the world. If we consider the inclusion of the spoken utterances in the animation scenes as the addition of one or two new attributes ("lexical label" and/or "verbal argument") our results show the same facilitation obtained by Kersten and Billman (1992) and predicted by the focused sampling procedure.

Language, however, may have a special status. Our syntactic distinctions did not just provide another cue that correlated with event structure. Rather, as English speakers, subjects had particular expectations about the relation between scene and syntax: they expected the entity that moved first to be the subject. The fact that all subjects made the same assignment between shape and syntactic argument filler and assigned referent of the noun accordingly provides evidence that they all shared a common bias. In the context of new language acquisition, systematic correlation of syntax would probably have a facilitating effect independently of the specific correlations. Additional comparisons would be needed to determine whether the same facilitation could have occurred with a different correlation between syntax and scene, one which disrupted either parameters of English or patterns of proposed universals.

The idea that language and thought interact, and interact in acquisition, is a central one throughout the cognitive sciences. Typically, researchers assume conceptual structure is fixed and ask how learning language draws on conceptual structure. The research begun here traces out one place where the reverse influence, from syntax to concepts, is felt. Analyzing the scope of effects from syntax to concepts is important for understanding acquisition in people, for developing natural language systems that learn, and for building computational models of human learning and understanding.

References

Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes, 4*, 127-155.

Billman, D. & Heit, E. (1988). Observational learning from internal feedback: A simulation of an adaptive learning method. *Cognitive Science, 12*, 587-625.

Billman, D. & Knudson, J. F. (1990). Systematic versus orthogonal sets of correlations benefit unsupervised concept learning. Paper presented at the *31st. Annual Meeting of the Psychonomics Society*. New Orleans.

Fisher, C., Gleitman, H. & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology, 23*, 331-392.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4*, 161-178.

Gleitman, L.R. (1990). Structural sources of verbal learning. *Language Acquisition, 1* (1), 3--55.

Grimshaw, J. (1981). Form, function, and the language acquisition device. In C. L. Baker and J. J. McCarthy (Eds.), *The Logical Problem of Language Acqisition.* Cambridge, MA: MIT Press.

Hunt, E. & Agnoli, F. (1991). The Whorfian hypothesis: a cognitive psychology perspective. *Psychological Review, 98*, 377-389.

Huttenlocher, J. & Lui, F. (1979). The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior, 18*, 141-162.

Huttenlocher, J., Smiley, P. & Ratner, H. (1983). What do words meaning reveal about conceptual development? In T. R. Seiler & W. Wannenmacher (Eds.), *Reader: Concept Development and the Development of Word Meaning*. Berlin, Germany: Springer-Verlag.

Kersten, A. & Billman, D. (1992). The role of correlational structure in learning event categories. *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Theory and Research* (pp. 64-99). New York: Appleton-Century Crofts.

Rosch, E. (1978). Principles of categorization. In Rosch & Lloyd (Eds.), *Cognition and Categorization.* Hillsdale: Lawrence Erlbaum.

Whorf, B. L. (1956). *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf.* T. B. Caroll (ed.). Cambridge, MA: MIT Press.