Psychological Review
1991, Vol. 98, No. 3, 377–389

0033-295X/91/$3.00

The Whorfian Hypothesis: A Cognitive Psychology Perspective

Earl Hunt
University of Washington

Franca Agnoli
University of Padova, Padova, Italy

The linguistic relativity (Whorfian) hypothesis states that language influences thought. In its strongest form, the hypothesis states that language controls both thought and perception. Several experiments have shown that this is false. The weaker form of the hypothesis, which states that language influences thought, has been held to be so vague that it is unprovable. The argument presented herein is that the weaker Whorfian hypothesis can be quantified and thus evaluated. Models of cognition developed after Whorf's day indicate ways in which thought can be influenced by cultural variations in the lexical, syntactical, semantic, and pragmatic aspects of language. Although much research remains to be done, there appears to be a great deal of truth to the linguistic relativity hypothesis. In many ways the language people speak is a guide to the language in which they think.

We talk about what we are thinking, and with occasional lapses we think about what we are saying. But does the language we use exert any control over thought? Do speakers of English, French, or Chinese think about common experiences in different ways? Eminent scholars have claimed that language does form thought. Herodotus believed that Greeks and Egyptians thought differently because Greeks wrote from left to right and Egyptians, from right to left (Fishman, 1980). Some millennia later Einstein (1954) wrote

> Thus we may conclude that the mental development of the individual, and his way of forming concepts, depend to a high degree upon language. This makes us realize to what extent the same language means the same mentality. In this sense, thinking and language are linked together. (p. 336)

The argument for linguistic control of thought was elegantly expressed by Benjamin Lee Whorf, an American businessman and amateur linguist who was active during the 1920s and 1930s. Whorf had studied with Edward Sapir, an anthropologist who held similarly strong views on the topic. Collectively, their views are known as the Sapir–Whorf (or Whorfian) hypothesis. Sapir and Whorf were modern spokesmen for a tradition established by the German Romantic philosophers of the nineteenth century. The German philosophers, in turn, had their predecessors. As our opening remarks about Herodotus indicate, this is a very old idea.

The cognitive revolution in psychology virtually rejected the Whorfian hypothesis. In their widely used textbook on psycholinguistics Clark and Clark (1977) stated

> What can one conclude about the Sapir–Whorf hypothesis? At present very little . . . languages can apparently be stretched and adopted to fit the needs of virtually any group of experts . . . language differences reflect the culture, and not the reverse. (p. 557)

Foss and Hakes (1978, p. 393) were somewhat more accommodating, allowing that there may be something to the Whorfian hypothesis but not in any form that Whorf himself would recognize. These conclusions are hardly surprising. Following Chomsky, the major trend in modern linguistics has been toward the study of linguistic universals, with a concomitant dismissal of the psychological importance of differences between languages. We will, however, note some exceptions.

The idea that language is a module in the service of thought is not compatible with the phenomenological experience of people who go back and forth from one language to another. Articulate bilinguals have maintained that they do think differently in different languages. Wierzbicka (1985b) said that she is a "different person" in Polish and English. Her further remarks make it clear that she believed that her attitudes and interpersonal behavior were affected by the language she was using. More succinctly, E. Hoffman (1989) titled her well-received book *Lost in Translation.*

We believe that this discrepancy between academics and bilinguals tells us something. It is time to take another look at the Whorfian hypothesis, making use of the progress that has been made in psychology and linguistics since Whorf's day.

Major Issues

Linguistic criticisms of Whorf have usually focussed on intertranslatability: Can a statement in one language be translated into a statement in another language (Lakoff, 1987)? The general consensus is that such translation is possible, although a concise statement in one language may map into a lengthy statement in another. The Kiriwina language of New Guinea contains the word *mokita,* which means "truth everybody knows but nobody speaks" (Rheingold, 1988). English-speaking politicians may not have the word, but they certainly have the concept.

Translatability contradicts the strongest version of the Whorfian hypothesis, which states that a thought expressible in one

We thank Megumi Kameyama for her discussion of Japanese examples and Christine Greenway for a description of the Quechua society and language.

Correspondence concerning this article should be addressed to Earl Hunt, Department of Psychology (NI-25), University of Washington, Seattle, Washington 98195, or to Franca Agnoli, Via Beato Pellegrino 26, Padova 35137, Italy.

language may not be expressible in another. The weaker form of the hypothesis states that language differentially favors some thought processes over others, to the point that a thought that is easily expressed in one language might virtually never be developed by speakers of another language. The issue becomes the degree of naturalness of thought. At the intuitive level, a thought is "natural" if it comes easily to the language user. Unfortunately, ideas like "natural" and "comes easily" are not scientific concepts.

But consider a computational theory in which language use depends on constructing propositions in working memory based on both abstract knowledge and language forms (e.g., Just & Carpenter, 1987). According to this view, a thought is natural if the necessary propositions can be constructed with a minimum of computation. The idea of naturalness becomes more specific if we consider the computational burdens involved in expressing different ideas in different languages.

The most obvious way to avoid computational burdens is to have a predetermined symbol structure identifying the thought to be transmitted. In mathematics this is established by notation. The symbol Σ transmits a concept to statisticians that takes several words to describe. Analogously, one language may have a single term for a concept that has to be described in another language. The mokita example previously given illustrates this principle.

The fact that languages differ in this way is not in itself a psychologically interesting distinction. Rosch (1974) correctly asserted that Filipino rice farmers and Eskimos talk about different things. Also, lexical additions are made to a language much more quickly than deep psychological processes could conceivably be changed. Evidence of this is seen in the use of words like *bulldozer* and *Boeing 747* in French and Japanese.

A more promising approach is to look at the number of decisions that a language user has to make to choose a word or to construct an utterance. In the mathematical theory of communication a message's complexity is defined by the number of decisions required to produce the message (Shannon, 1948). In more detailed analyses, the computational complexity of a message is defined by the uncertainty of the individual components and the correlation between these components. Consider how these concepts apply to the formation of a linguistic message.

Imagine a speaker who wishes to express a proposition. The problem is to choose a lexical unit for each term in the proposition. From the viewpoint of computational theory this problem will be affected by the choices available from and required by the lexicon.

Pronouns provide a concrete example. In English the selection of a pronoun is largely driven by biology, although a speaker does have to know certain conventions, such as the custom of referring to ships as "she." Chinese pronouns are not marked for gender so no decision is required to select the pronoun that refers to Mary or John. In German the decision-making process is even more complex because the biological category of sex and the syntactical category of gender are loosely correlated and may even conflict. Mark Twain (1880/1977) observed with exasperation that *mädchen* (girl) is neuter but *Rübe* (turnip) is feminine.

Coding considerations determine the demands that a language places on its users' psychological capacities. Recognizing and selecting lexical items places demands on long-term memory. Analyzing the structure of an utterance taxes short-term memory. The historical record suggests that languages evolve to move the burden from the short-term to the long-term memory system. Hunt and Banaji (1988) observed that in the last 20 years southern Californian surfers have invented a vocabulary for describing waves, which includes the descriptive terms *hollow* and *flat*. Presumably the surfers of the 1950s could describe the same waves by using sentences, thus increasing the burden on their short-term memory. The modern surfer has traded expensive space in short-term memory for cheaper space in long-term memory. A similar process has been observed when people learn to be skilled in such specialized tasks as remembering digits (Ericsson, 1985). This means that at any point in time a language user thinks most efficiently about those topics for which his or her lexicon has provided an efficient code.

The complexity argument may be actually closer to Whorf's own ideas than the extremely stated Whorfian hypothesis. He believed that the distinctions that are made implicitly, by the grammar of a language, are far more important determinants of thought than are the explicit categorizations of the lexicon (Whorf, 1956, pp. 80, 90–92).

Whorf's critics have argued that the complexity argument is not useful because it cannot be tested. (Foss & Hakes's, 1978, criticisms are essentially of this nature.) The negation of the complexity argument is that language does not affect thought and this, being a version of the null hypothesis, cannot be proven. The complexity argument can be tested if we choose some minimal effect size and ask whether there are Whorfian effects that are at least as large as this minimal effect. But what is an appropriate size for a minimal effect?

The only way that we can think of to determine a minimally interesting effect size is to examine current practice. For example, psychologists and psycholinguists are willing to study small differences in the time required to comprehend a statement. Consider the effect of markedness on comprehension. It is very well established that it takes longer to comprehend a sentence like "The plus is below the star" than it does to comprehend "The plus is above the star." The absolute difference in comprehension times is about 50 ms. Because this finding is included in many texts on psycholinguistics, cognitive scientists must regard a 50-ms effect as significant in the absolute rather than the statistical sense. Similarly, cognitive psychologists regularly (and in our belief properly) report effects that account for about 10% of the variance in a variety of measures of linguistic performance.

These observations suggest a way to reformulate the Whorfian hypothesis. Can we find crosslinguistic effects that are as large as the intralanguage effects cognitive scientists are willing to study? This formulation of the Whorfian hypothesis can be tested empirically.

We develop our argument in a structured manner. First we present a brief statement of the view of cognition that is found in many discussions of modern cognitive psychology (e.g., Anderson, 1983; Hunt & Lansman, 1986; Klahr, Langley, & Neches, 1987). We then present illustrations of interactions between language and cognition that are suggested by this view. We separately discuss lexical effects, utterance-level effects (i.e., syntax and semantics related to the analysis of well-formed ex-

pressions), and higher level effects, including the relation between language, schema-based reasoning, and the analysis of the pragmatic meaning of utterances. For each class of effects we present examples that support the Whorfian hypothesis, examples that question it, and suggestions for studies that might resolve the issues involved.

At times our arguments may strike psychologists as somewhat unusual. Although we primarily rely on the methods of experimental psychology, we shall not hesitate to report field observations, propose thought experiments, or, on occasion argue from intuition. We believe that this is necessary because language is too complex to be contained within the paradigm of any one science.

A View of Cognition

Pylyshyn (1984) has pointed out that there are three levels at which cognition can be studied. The lowest level is concerned with the physiological mechanisms underlying thought. These are presumably cultural universals and do not concern us. The highest level is concerned with the content of thought; that is, what aspects of the world are coded in the mind? Following the lead of Holland, Holyoak, Nisbett, and Thagard (1986), we refer to this as the representational level of theorizing. Between the representational level and the physiological level are the "mechanics" of how a representation is formed without regard to the content of that representation. Newell (1990) has referred to this as the symbol manipulation level of thought. Elsewhere, with more particular reference to verbal comprehension, Hunt (1978, 1980; Hunt & Poltrock, 1974) has used the term *mechanics of thought*. We first consider the mechanical–symbolic level and then the representational level.

We assume that when language stimuli are received they are converted from a visual or auditory code to an abstract lexical code. This implies the existence of a lexical access system that can connect a physical symbol to the lexical entries it might represent. Once lexical items are accessed the semantic and syntactical relations associated with them are used to form propositional statements in working memory. Some of the propositional statements are then incorporated into a long-term memory structure that loosely represents the comprehender's idea of what is going on.

The previous paragraph seems to say that the comprehension system is completely stimulus driven. In fact no one believes this. We see all stages of comprehension as a process operating in a "blackboard model" (Erman, Hayes-Roth, Lesser, & Reddy, 1980; Hayes-Roth & Hayes-Roth, 1979) in which recognition of a lexical item is facilitated by prior recognition of related items and propositional statements, or both. Similarly, propositional analysis of the current linguistic utterance is influenced by the structure of the discourse model that exists when the propositional analysis takes place. See Just and Carpenter (1987), Hunt and Lansman (1986), and van Dijk and Kintsch (1983) for discussions of how language analysis in general can be achieved by such a model.

In language processing mechanical effects influence the efficiency of construction and manipulation of propositions as symbol structures without regard to the meaning of those structures. The sentences "John loves Mary" and "Pigs have wings" are presumably processed in the same way at the mechanical level even though their representational level meanings are distinct.

In contrast, representational level thinking is concerned with meaning. The product of linguistic analysis is a symbol structure that represents a nonlinguistic reality. To develop this representation a comprehender will use any information that he or she has. Two sources of information are particularly important to the Sapir–Whorf hypothesis: Lexically identified concepts that serve as the primitive elements of the discourse structure and culturally developed schema that serve as blueprints to guide the comprehender in building the structure.

To illustrate by analogy, suppose that a person is trying to assemble a mail-order computer. In the 1960s this would have been a difficult task because the person would have received a collection of elementary logic circuits to be arranged into such things as arithmetic and memory units. In the 1990s the kit would include the memory and central processing unit on small chips, and assembly would not be terribly difficult. Similarly, most feasible mail-order projects are accompanied by a blueprint that tells the builder how to proceed. In most cases a blueprint does not specify the only method of construction, but it does specify a feasible one. Also, instructions vary greatly in their specificity. Sometimes there is only one interpretation possible, and other times the instructions restrict, but do not dictate, a builder's choice of actions.

This is a fairly precise analogy between model building and the representational analysis of language. The comprehender is trying to construct a model of what is in the speaker's mind (Sperber & Wilson, 1986). Consider a statistics student listening to a lecture on the Poisson distribution. The student's problem is to construct a discourse-level model of the lecture. This can be done only if the student receives manageable linguistic messages from the instructor. If the instructor uses the term *Bernoulli sequence,* that term has to identify a concept (building block) that the student already has. If the instructor has to develop the concept of a Bernoulli sequence as an aside, the main point of the lecture can easily be lost.

We argue that different languages lend themselves to the transmission of different types of messages. People consider the costs of computation when they reason about a topic. The language that they use will partly determine those costs. In this sense, language does influence cognition.

Lexical Influences

In the opening section we argued that differences in lexicons are not psychologically interesting just because they exist. Differences in lexicons become psychologically interesting when they cause speakers of two different languages to structure the same experience in different ways. We will consider two classes of lexical effects. The first we call *direct effects* because they depend on discriminations that a person must make in choosing or comprehending a word. The second we call *indirect effects* because they depend on the semantic relations between the word chosen and other words in the speaker's lexicon.

When a word is chosen it focuses attention on a particular aspect of experience that makes the word appropriate. Indeed, speakers intend that this should happen when they choose

words. "Here comes the watchdog" is different from "Here comes the family pet." Whorf noticed that Navajo words emphasize the structure of objects, whereas English words do not. He then conjectured that the Navajo speaker paid more attention to form than the English speaker did. Carroll and Casagrande (1958) attempted to experimentally validate Whorf's conjecture. Subsequently their study has been cited both as evidence for and against the Whorfian hypothesis. Therefore it is worth explaining here what Carroll and Casagrande did and what they found.

Carroll and Casagrande (1958) examined the classification performance of rural Navajo children who were dominant in either the Navajo or English languages. The children were shown a pair of objects that varied in size and form, for example, a yellow rope and a blue stick. They were then given a comparison object that matched each of the items in the pair on one dimension (e.g., a blue rope). The children's task was to say which item belonged with the comparison object. Navajo-dominant children chose on the basis of form about 70% of the time. English-dominant children chose on the basis of form about 40% of the time. The data were not so striking for a form–size contrast but they were in the same direction. Carroll and Casagrande concluded that these results support the Whorfian hypothesis.

Their conclusion was weakened by another aspect of the study. Carroll and Casagrande (1958) also examined the classifications of White American middle-class children from the Boston area. These children, who presumably had never heard a word of Navajo, chose form over color about 80% of the time (i.e., they exhibited a stronger "Navajo language bias" than the Navajo did). This result has lead the opponents of Whorf to cite Carroll and Casagrande's work as evidence for their side of the argument.

The argument is over the right control group. Carroll and Casagrande (1958) said that the English-dominant Native Americans were the appropriate control group because they were more similar in culture to the Navajo-dominant Native Americans than were the White Bostonians (see also Carroll, 1963). This argument is weak but it cannot be dismissed out of hand. Our conclusion is that, in spite of the importance that the Carroll and Casagrande result has acquired in the literature, the results were so inconsistent that the study should be redone.

It would not be necessary to redo this study with Navajo subjects (whose own language status is changing). Other linguistic groups are available. The Japanese language uses a system of classifiers that specify object shape. The sentence "Give me three pens" would be stated literally as "Give me three long and thin objects of pens." The Carroll and Casagrande study could be repeated using Japanese subjects.

With modern multidimensional scaling techniques a much more sophisticated study could be done. Stimuli could be designed to vary in both form and color. Participants could then be asked to rate similarity. The speaker's semantic space could then be derived using standard multidimensional scaling techniques (Luce & Krumhansl, 1988). Would Japanese and English speakers differ in the extent to which they placed weight on color or form? Would the semantic space of bilinguals depend on the language in which they were examined?

These questions are directed at the influence of language on the relative perceptions of two different dimensions of stimulus variation. Language effects may also be evident within a single dimension. Historically this question has been investigated by studying color. Different languages make different semantic distinctions between hues. Ancient Hebrew had a single term for the two colors English speakers call green and yellow. Classic Greek did not distinguish between blue and black. Italian has four frequently used terms (*blu, celeste, azzurro,* and *turchino*) corresponding to the single term *blue* in English. The Japanese term *awo* can mean green or blue depending on context. Do the different naming patterns have any psychological significance? Do Italians see finer distinctions between shades of blue than do English speakers?

The question is not quite this straightforward. The visual system represents a hue as a mixture of three focal colors corresponding to, in English, the colors called red, green, and blue (Caelli, 1981). On physiological grounds we might expect these colors to have special status. Furthermore, languages do not assign color names haphazardly across the spectrum. Berlin and Kay (1969) showed that there are at most 11 basic color terms and that they are assigned in an orderly hierarchy. There are a few languages that only have terms for black and white. If a third color term is added it is red. The next three terms added are yellow, blue, and green. Across languages speakers may differ in defining the boundaries between two colors but will agree on the best (prototypical) example of any of the categories that the two languages share. Berlin and Kay's analysis has been interpreted as showing that there is a universal, physiologically based principle behind color naming.

This interpretation was supported by Rosch's (1973; Heider & Olivier, 1972) well-known study of color perception and memory in the Dani, an aboriginal people in what is now Papua New Guinea. The Dani language has two color labels, black and white. When Dani speakers were shown a color chip and then tested for recognition of that chip a few seconds later, they performed better if the chip was a focal color than if it was not. Rosch then taught Dani participants arbitrary associations between nonsense syllable names and color chips. Chip names were learned more easily for focal colors than nonfocal colors. This work has been cited widely as providing strong evidence against the Sapir–Whorf position (Clark & Clark, 1977; Foss & Hakes, 1978).

We disagree for two reasons. First, recent findings by anthropologists limit Rosch's conclusion. Second, we do not believe that color perception is as crucial a test of the Whorfian hypothesis as Rosch and others have presented it to be.

Lucy and Shweder (1979) conducted an experiment on color discrimination using the methods of triads. Participants were shown three chips, two of the same color and one that differed very slightly from the other two. The task was to select the two same-colored chips. English speakers were able to discriminate focal colors better than nonfocal colors. The result, however, could not depend on language because the task does not involve labeling. This finding suggests that Rosch's focal color chips were simply more distinctive as perceptual events than were nonfocal chips.

Lucy and Shweder (1979) also had English speakers participate in a naming game in which one person had to describe a

chip to a second person who would then identify the chip being described. The chips were chosen from chips that had been found equally discriminable in the perceptual experiment. Lucy and Shweder analyzed the communications between participants to develop an index of linguistic codability for each chip. Linguistic codability was related to the ability to recognize chips in a memory experiment. The relation was slightly stronger if the memory interval was increased to several minutes instead of restricting memory to the 30-s interval used by Rosch.

Lucy and Shweder (1979) pointed out that their results were consistent with previous anthropological studies of color memory in Mayan and Spanish speakers (Lantz & Stefflre, 1964; Stefflre, Castillo Vales, & Morley, 1966). Some hues have different codabilities in Mayan and Spanish. When shown the same chips Mayan and Spanish speakers differed in their memory capacities. The difference was related to color codability in the speaker's language.

These studies show that although *perception* may be relatively immune to language, *memory* is not. Memory can be based on two different records, a direct record of the sensory information at the time that we perceive an event and an indirect, linguistically based record of our description of the event to ourselves. The latter effects, because they are coded by language, are subject to any biases built into the memorizer's language.

Such effects have been shown many times in studies of eyewitness memory where recall and recognition can be biased by the experimenter's introducing verbal information that alters the verbal record. For instance, in a widely cited study Loftus and Palmer (1974) showed people a green car. The car was then labeled *blue* by the experimenter. In a subsequent recognition test observers chose as the true color one that was intermediate between the color originally presented and a prototypical blue color. Schooler and Engstler-Schooler (1990) extended this result to names generated by the observer. People were shown nonprototypical color chips, which they were then asked to label. A control group saw the same chips but were not asked to label them. The experimental group performed worse than the control group on a subsequent recognition task, indicating that the act of labeling distorted memory. Because this research was done in the context of eyewitness memory, Schooler and Engstler-Schooler did not comment on the importance of their study for the Whorfian hypothesis. We feel that it is relevant because labeling has to be restricted to terms in the observer's language. Therefore the extent of distortion should be determined by the fineness of the labeling.

In fact, precisely this sort of finding has been reported. Kay and Kempton (1984) created triads of colors in which two items (A and B) were clear examples of blue and green and the third member of the triad (C) lay between them. People were then asked to decide whether the C chip was closer to the A or B chip. English-speaking subjects showed categorial perception in the sense that they sharply distinguished between chips that lay on one or another side of the English color boundary. People who spoke Tarahumara, a Mexican-Indian language that does not have blue and green color terms, did not show categorial perception. Thus the naming strategy exaggerated the perceived difference between two hues. In a subsequent experiment Kay and Kempton showed that the categorial effect could be destroyed if English speakers were induced to call the intermediate chip both blue and green, thus eliminating its association with just one name.

In closing our discussion of direct effects we would like to make what we think is an overlooked point. Kay and Kempton (1984) illustrated a language effect on color perception by demonstrating variations in categorial perception across languages. Categorial perception had not been discovered in Whorf's day. Categorial perception is, however, a striking example of how language can control perception, albeit at the phonetic rather than the lexical level. English speakers show categorial perception for sounds varying continuously from a pure /l/ to a pure /r/. Japanese speakers, whose language does not make the /l/ versus /r/ distinction, do not show categorial perception (Miyawaki et al., 1975). This is a confirmation of the strong version of the Whorfian hypothesis.

Now we move from direct to indirect effects. These are lexical effects produced by relationships between words, rather than effects that are produced by the semantics of an individual word. Unintended and uncontrolled activations of different meanings may vary systematically across languages. In addition, languages may vary in the semantic hooks that a word provides to connect it to other words in an utterance. Here we want to distinguish between semantic meanings that are used to identify a word's function in an utterance, which will be discussed in the section on syntax, and cues that can be used as heuristics to assist in the determination of the meaning of a discourse. To illustrate, the syntactic constructions of the sentences (1) "John shot the moose" and (2) "John shot the picture" are identical. However the semantics of "shot" prepare us for sentence (1) much more than (2) even though (2) is probably statistically more likely. The question is, do different languages systematically do a better or a worse job of such preparation? And if so, to what extent do these differences influence utterance comprehension?

Polysemous words provide a particularly interesting example of indirect effects. Current theoretical and experimental studies of priming have shown that when a polysemous word is encountered all its meanings are activated. The correct meaning is then selected by an analysis of context (Ratcliff & McKoon, 1988). Consider two languages, both of which have the same term for referent *A*. Suppose that in the first language term *A* has a polysemous secondary meaning *B*, but in the second language the secondary meaning is *C*. When the speakers of these two languages refer to *A*, will one speaker activate referent *B* and the other referent *C*?

The most extreme case of such an effect would be when the first language has two references and the second language has only one (i.e., when the contrast is between an *A*, *B* reference and an *A* reference). To illustrate this we compared English and Italian. Thirty English words and 30 Italian words were selected at random from *Il Nuovo Ragazzini* (Zanichelli, 1984), a 2,112-page Italian–English dictionary with 120,000 entries. For English words the mean number of meanings was 5.67 and the standard error was 1.22. For Italian words the mean number of meanings and the standard error were 2.83 and .52, respectively. English words were significantly more polysemous,

$t(58) = 2.11$.[1] We then repeated the study using the *Random House Italian–English Dictionary* (with 30,000 entries) that is restricted to high-frequency words. Sixty English and 60 Italian words were randomly selected from this dictionary. We then compared the mean number of meanings as they appeared in the *Nuovo Ragazzini* dictionary. The statistics were virtually identical to those in the original study.

Do differences in the polysemy of languages have any cognitive effects? When a sight or sound is identified as a lexical entry the comprehender still has to determine the appropriate semantics. For nonpolysemous words the two steps are one. In a language like English (which seems to have carried punning to an art form of sorts) identifying the lexical entry is only the beginning of processing. The sentence (3) "I went out to buy the pot" can only be disambiguated if we know whether the speaker spends leisure time in gardening or recreational pharmacology. Perfetti and his colleagues (Perfetti, Beverly, Bell, Rodger, & Faux, 1987) showed that such sentences require demonstrably more time to process than do unambiguous sentences, even when the meaning is clear from context.

Note how this would affect the intertranslatability of sentences like (3). For a gardening pot an Italian would say (4) "*Uscii a comperare il vaso*," which does not have an alternative meaning in Italian. Presumably, Perfetti's effects could be demonstrated with (3) in English but not with (4) in Italian.[2]

Polysemy exerts an influence by directing attention to semantic dimensions that are irrelevant to the present discourse. Languages also differ in the extent to which they use words to direct attention to relevant semantic dimensions. In addition to the use of different contrast classes across language (see the previous discussion of form), languages use different semantic distinctions as devices to convey syntactical and intrautterance referential information. This can be illustrated by anaphoric reference.

An anaphoric reference establishes an identity between the meaning of a term at the reference's location and a previously introduced topic. A language's anaphoric rules can call attention to particular aspects of the referent. In English the pronomial referents *he* and *she* call attention to gender assignment, which depends on semantic knowledge that may or may not be equated with world knowledge. When referring to animals English speakers will normally say "it" unless there is some reason for stressing the biological sex of the particular animal. Italian and Spanish speakers must make a discrimination based on syntactic gender. To illustrate, when referring to a seal an English speaker will say "it" unless the sex of the animal is known. The Italian will say *essa* (she) because *la foca* (the seal) has the syntactic feminine gender. It is an open question whether or not forced references to gender have an influence on our thinking about those objects. Of course, we do not mean to imply that Italians are unaware of the fact that wolves (*lupo*, masculine) and seals (*foca*, feminine) have male and female categories. We do wonder about the extent to which syntactic gender assignments influence mental representations of the relevant object. Consider the following three thought experiments.

English speakers locate animals in a semantic space whose principle dimensions are size and domesticity–predacity (Henley, 1969; Rips, Shoben, & Smith, 1973). Do Italian and Spanish speakers do the same? Is there any correlation between the location of animals in the semantic space and a language's assignment of syntactic gender?

In release-from-proactive-inhibition experiments people are given three or four words to memorize, engaged in a distracting activity for a few seconds (Peterson & Peterson, 1959), and then asked to recall the words. On the first trial of such an experiment retention is virtually perfect. On subsequent trials recall deteriorates, providing that the items to be recalled are taken from the same semantic category on each trial. If semantic categories are switched (e.g., from men's names to women's names in English) retention is again excellent. However, switching syntactic categories (e.g., nouns to verbs) does not produce release from proactive inhibition for English speakers (Wickens, 1972). Does the Italian–Spanish gender dimension act like the English semantic or syntactic distinction? Notice that if proactive inhibition were shown to be controlled by different distinctions in different languages this would be a pure example of the Whorfian hypothesis. A cognitive-function, short-term memory would be shown to be under linguistic control.

Suppose that this experiment had positive results. An anti-Whorfian might argue that even though the effect had been shown, it was shown only in the context of an artificial laboratory task. Our next thought experiment deals with short-term episodic memory in a more natural linguistic situation.

Different languages force us to describe events in different ways, in the sense that they focus attention on different aspects of the linguistic situation. Does the description enforced by the language alter recall of the event? Suppose *A* tells *B* the following tale: (5) "I was scared because I saw a bear with her cubs. I ran before she saw me." In this English sentence the sex of the bear has been stated three separate times, twice through the use of the pronoun and once by context; only female bears accompany cubs. These forms are compulsory.

In Italian this tale could be written as follows: (6) "*Presi paura perché vidi un orso con i suoi piccoli. Scappai prima che mi vedesse.*" Translated literally these sentences state "Was scared because saw a bear (masculine) with the belong-to-subject-of-preposition cubs. Ran before that me saw." Here the sex of the bear is inferable only from context. Would an English speaker who has been told about the bear three times have better memory for the animal's biology?[3]

---

[1] This difference can be partly explained by a well-known attribute of English—that many words can serve as either a noun or a verb (*a cry, to cry*). In Italian this form of polysemy is rare, even though nouns often take the same form as verbs in the present form of the indicative mood, for example, *un grido* (a scream) or *io grido* (I scream). To assess whether this accounts for the entire difference in English and Italian polysemy, we examined the polysemy of only noun interpretations. For English the sample included 29 nouns with a mean and standard error of 3.14 and .71 meanings, respectively. For Italian, the sample included 23 nouns with a mean and standard error of 1.91 and .23, respectively. English words were more polysemous, but the difference was not significant, $t(50) = 1.48$.

[2] Interestingly, though, polysemous words presented in isolation are identified faster than nonpolysemous words.

[3] An Italian could make the distinction. It would be possible to use *orsa*, a low frequency term roughly equivalent to the English term *she-bear*, or to say *lei mi vedesse* (she saw me), or both. The point is that the English speaker must transmit the information, but the Italian must make a conscious decision to do so.

As our examples indicate we are moving from a discussion of lexical effects to a discussion of utterance-level effects. The next section goes into this topic in greater detail.

Syntactical–Semantic Effects (Utterance Analysis)

Much of the power of human language lies in the ability to construct utterances. An utterance is a set of lexical items that conform to certain rules for well-formed structure. Note that we said *set* rather than *string*. The change was intended to stress what we believe is an important point. In English word order is the major cue to syntactical structure. That is how we distinguish between the prosaic (7a) "dog bites man" and the newsworthy (7b) "Man bites dog." Classic Latin made the distinction by morphology instead of word order. In Latin (b) would be expressed as (8a) "*Vir canem mordet*," literally "Man (nominative) dog (accusative) bites" or perhaps as (8b) "*Canem vir mordet*." If *vir* is substituted by *virum* and *canem* by *canis* then the dog bites the man regardless of word order.

English speakers rely heavily on their memory for order. The importance of this ability has been shown by studies of individual differences in verbal competence. English-speaking students who have better than average verbal comprehension also have better than average memory for order information, although they do not have unusually good memory for item information (Hunt, Frost, & Lunneborg, 1973). This finding has been presented as a fact about "high verbals." It may be a fact about "High English." Would the same correlation between verbal competence and memory for order be found in an inflected language such as Spanish or Italian?

Accurate memory for order information seldom extends beyond seven or eight items. Comprehensible English sentences may be much longer than this. English speakers cope with the problem by developing strategies for attaching lexical items to their syntactical position as the lexical items occur (Just & Carpenter, 1987). An example of such a strategy is late closure (Frazier, 1987), a rule that states, "Attach new items into the . . . phrase or clause postulated most recently." Late closure helps to minimize the number of items that must be held in memory until their syntactical location can be determined without error. The strategy usually works but can create errors when people hear "garden path" sentences, such as "The horse raced past the barn fell."

Such sentences force people to reanalyze their recall of the initial words in an utterance. Not surprisingly, individuals with larger working memories have an easier time recovering the meaning of a garden path sentence than do individuals with smaller working memories (Daneman & Carpenter, 1983).

The late closure strategy makes sense for analyzing English but the argument for late closure is less compelling in an inflected language, where a requirement for order memory is weaker. There is evidence that the late closure strategy is not as prevalent in Spanish or French as it is in English (Cuetos & Mitchell, 1988; Mitchell, Cuetos, & Zagar, 1990). When presented with the sentence "Andrew had dinner yesterday with the niece of the teacher who belonged to the Communist party" "*Andrés cenó ayer con la sobrina del maestro que está en el partido comunista*", 58% of English speakers judged that the teacher belonged to the Communist party (late closure) but only 37% of the Spanish informants did. Similar results were obtained with French speakers, showing evidence against a preference for late closure in two inflected languages (Mitchell et al., 1990).

These results suggest that different languages use different cues for parsing. Extensive evidence to support this statement comes from data collected by Bates and MacWhinney and their colleagues (MacWhinney & Bates, 1989). They conducted experiments on sentence comprehension in different languages. The standard experimental task was to present native speakers with a series of simple transitive sentences composed of two nouns and a transitive verb, as in "Camel pencil lick." Native speakers were asked to decide which of the two nouns was the subject. The linguistic cues varied from experiment to experiment; however, they always included a set of the following: different levels of word order (noun–verb–noun, noun–noun–verb, verb–noun–noun), different levels of animacy (animate vs. inanimate nouns), grammatical morphology (e.g., subject–verb agreement, agreement between object and clitic pronouns), prosodic contrasts, and, in some cases, different forms of topicalization.

English, Italian, and German native speakers showed wide variations in their use of different cues (Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; MacWhinney, Bates, & Kliegl, 1984). English adults relied largely on word order. Subject–verb agreement and animacy were almost completely disregarded. In Italian, on the contrary, word order had little effect. Subject–verb agreement and animacy were the strongest cues to sentence comprehension.

This difference between Italian and English comprehension strategies is compatible with the differences in the two languages. English morphology is impoverished compared to Italian. There are fewer permissible word-order variations in English than in Italian.

German speakers' selections of subjects were determined by agreement and noun animacy cues in accordance with the rich German morphological system, which provides a regular set of markings for tense, number agreements, and case, (see also Kilborn, 1989). Similar contrasts have been obtained when the experiments were repeated in other languages (see Kail, 1989, for French and Spanish; Kilborn & Ito, 1989, for Japanese; MacWhinney, Pléh, & Bates, 1985, and McDonald, 1987a, 1987b, for Dutch; Miao, 1981, for Chinese; Pléh, 1989, for Hungarian; Sokolov, 1989, for Hebrew).

Additional empirical evidence for the differential strength of cues in different languages can be found in studies of bilinguals tested in their first and second languages. Learners transfer their first-language sentence-processing strategies to sentence processing in the second language. The influence of this transfer can be detected in weakened form even in fluent bilinguals who have spoken the second language for many years (see Gass, 1987, for Italian–English and English–Italian; Harrington, 1987, for Japanese–English and English–Japanese; McDonald, 1987a, 1987b, for English–Dutch and Dutch–English; Miao, 1981, for English–Chinese bilinguals; Wulfeck, Juarez, Bates, & Kilborn, 1986, for Spanish–English).

These examples justify our distinction between the structure of a set of words and the structure of a string of words. Most textbooks on linguistics present syntax as a set of rules estab-

lishing the appropriate order of lexical terms in a string. But most linguists are English speakers. The use of word order as a major cue to syntax, in general, may be a feature of the aberrant English language.

These distinctions would be of little interest to a Whorfian if they had no implications for cognition. But we think that such implications exist. Our reasoning is straightforward. To use language a person has to be able to figure out what an utterance means. This can be done entirely within the linguistic system or it can be done by moving outside the linguistic system to determine what meaning is plausible. Therefore, if languages differ in ambiguity they force different styles of reasoning on their speakers.

We maintain that, in fact, languages do differ in ambiguity. In the most obvious case polysemy produces ambiguity, and we have already provided evidence that languages differ in polysemy.

Similar ambiguities can occur in parsing. One of us once observed a French road sign that said simply "*Enfants dangereux*," literally "Children dangerous." Does this mean "Watch out for the children" or "WATCH OUT for the children"?

Although this example shows that syntactical ambiguity can occur in an inflected language, it seems more likely to occur in a language with restricted inflection. The American newspaper headline "Police can't stop drinking" has at least three meanings in English. The Italian translations[4] of the two dominant ones are "*La polizia non può trattenersi dal bere*" ("The police cannot stop themselves from drinking") or "*La polizia non può trattenere il bere*" ("The police cannot stop the drinking"), both of which are unambiguous. (We leave the detection and translation of the third meaning as an exercise for the reader.)

Aaronson and Ferres (1986) suggested that context sensitivity is even more important in Chinese than in English. Kameyama (1987) makes the same case for Japanese. Let ∅ be the zero pronominal symbol. The utterance "*∅ Mary ga suki desu*," in isolation, can mean either "(someone) likes Mary" or "Mary likes (something/someone)." In fact, though, Japanese speakers would not normally find the sentence ambiguous because their language contains overt encodings of notions such as topic and point of view. Once these are known the meaning of the previous quote is clear.

Is it the case that the utterances that are actually spoken tend to be more ambiguous in Chinese and Japanese than in English and in English than in Italian? We do not know, but this is an empirical question that could be answered by appropriate research. To return to a point we introduced in the beginning of this article, it is obvious that no language has ambiguities that render communication impossible. The size of the information processing burden in each language is an open question.

Let us be more specific about the sort of empirical questions that could be generated. If two languages are relatively high in polysemy, the more words that are used in a sentence the more chance there is for ambiguity. Because word order is a more restricted device for binding than inflexion, longer sentences should be more prone to ambiguity in a word-order language than in an inflected language. We can then ask two questions. First, are these statements true? Second, do speakers of the language adopt their utterance-forming habits to avoid ambiguity? And if they do, to what extent do these habits influence their understanding of utterances?

We are confident that at least two languages differ in utterance ambiguity. As a small experiment, two Italian and two English speakers, comparable in formal education, read passages on the same topic (the 1989 movement of East Germans to West Germany) published by similar Italian and American newspapers. The English speakers agreed that 18 of 33 sentences were potentially ambiguous out of context. The Italian speakers found that only 3 of 64 sentences were ambiguous.

We believe that this striking difference is due to the contrast between English (word order and relatively high polysemy) and Italian (inflected and relatively low polysemy). As a point for future research, one could plot the number of interpretations of a sentence as a function of sentence length. If we are correct, the number of utterance interpretations per word (the slope of the function) should be higher in English than in Italian. However, if speakers have adapted to the deficiencies of English, the average English sentence should be shorter than the average Italian sentence.

This appears to be the case. Bates and Devescovi (1989) asked English and Italian speakers to describe visual scenes. Italian speakers produced more complex sentences than English speakers. Bates and Devescovi suggested that the Italians' use of morphological inflections allowed them to produce relative clauses that would have confused the English speaker.

The previous examples show how languages impose different cognitive burdens on their speakers. We could argue that even if this is true it is irrelevant to the Whorfian hypothesis, because the hypothesis asserts that languages affect how their speakers think about the nonlinguistic world, although we have shown that different languages propose different challenges to the cognitive system.

There are two answers to this argument. One is that linguistic reasoning often occurs concurrently with nonlinguistic reasoning and that the complexity of linguistic analysis will affect concurrent nonlinguistic thought (Yee, Hunt, & Pellegrino, in press). The other answer shifts the discussion from language performance to language learning.

Children learn to use whatever linguistic cues their language provides them. The "standard" linguistic argument is that the learning of syntax amounts to setting parameters in a universal grammar (Pinker, 1990). If this were true, we would expect little interaction between language learning and the way in which a child structures his or her environment. However there is an alternative position.

Slobin (1985) proposed a set of "operating principles" to describe first language learning. The gist of his idea is that children have an innate capacity both for structuring linguistic stimuli and for structuring the extralinguistic world (the *semantic space*, in Slobin's terms). The operating principles are rules that can be used to establish correspondences between selected

[4] The Italian translations here sound forced to a native Italian speaker. A more appropriate translation is to paraphrase, as in *La polizia non può impedire che la gente beva*, literally, "The police cannot stop that people drink," which makes the subordinate clause explicit.

features of the semantic and linguistic spaces. Slobin pointed out that during this process the child will learn those dimensions of the semantic space that are related to linguistic variation in his or her language. In commenting on Slobin's ideas, Bowerman (1985) remarked that such learning singles out some aspects of the nonlinguistic world as being more important than others.

Subsequently, Slobin (1989; Slobin & Bocaz, 1988) offered a specific case of how this might work: The contrast between the ways in which Spanish- and English-speaking children describe past events. The English-speaking child emphasizes action, for example, saying "The boy climbed a tree." The Spanish child emphasizes the result of the action, "*El niño está subido arriba de un árbol*," literally, "The child is climbed in the tree" (i.e., the boy is in a state of having climbed the tree). Although this example is compelling, as experimental psychologists we would like to see a complementary psychological experiment showing that in fact English children had paid more attention to the action, whereas Spanish children paid more attention to the result.

Slobin's model of language learning is not the only model that predicts Whorfian effects. Bates and MacWhinney (1987, 1989) presented a theoretical model of language learning, the competition model, that also predicts that language learning will alter a child's sensitivity to nonlinguistic features. The competition model is a loosely stated connectionist model in which language forms are probabilistically related to semantic functions. Bates and MacWhinney argue that a connectionist model of language learning must include "hidden units" that intervene between nonlinguistic input and linguistic output. The hidden units can be thought of as internal responses to configurations of primitive semantic cues. That is, they are internally generated stimuli that are surprisingly like the "response produced stimuli" that appeared in behaviorist accounts of cognition in the 1950s and 1960s. Once such cues are established they can be used as building blocks in further learning. Thus, Bowerman's comment about how language learning sharpens the salience of particular semantic aspects of the world, although originally intended as a comment on Slobin's model, also applies to Bates and MacWhinney's connectionist model.

Representational Level Issues: Schemas and Pragmatic Meaning

The topics we have raised so far revolve around low-level effects of language. One of us (Hunt, 1978) has called such effects the *mechanics* of language comprehension. We suspect that many people (possibly including Whorf and certainly including Sapir) would object that the Whorfian hypothesis is not about mechanics. The more sweeping versions of the Whorfian hypothesis assert that language affects our interpretations of the world around us, not the size of the information-processing burden we have to overcome to reach those interpretations.

Milliseconds are not the right metric for measuring Whorfian effects on comprehension. It is more appropriate to study how language influences the schema used to order nonlinguistic experiences. This notion requires some explanation.

A schema is a set of directions that tell an observer where and when to look for information, what to expect to find, what to assume when a specific piece of information is missing (default values), and how and why to make inferences from the information one receives. We find it useful to distinguish between highly restricted schema, which provide detailed expectations about a small part of the world, and nonrestricted schema, which provide looser expectations about a wider range of variables.

Number schema are prototypical examples of restricted schema. They provide a way of discussing abstract properties of sets of objects. For instance, the number schema used in the languages of industrial societies provide a complete mapping between linguistic terms and the cardinality of sets. English speakers have no difficulty expressing the idea that, if there are 49 men and 37 pairs of shoes, some men will have to go without shoes. There are nonliterate societies where this would be a difficult situation to describe, because the language may have number terms only for "one–two–many" (Greenberg, 1978). This example entails more than just vocabulary. Number schema are essential parts of arithmetic, and there are many acts of nonlinguistic cognition that require arithmetic concepts.

A particularly interesting example is provided by languages that utilize body analogies to define numbers. Saxe and Posner (1983, p. 310) provided a striking example of a Whorfian effect that is due to such languages. Children who speak Oksapimin, a New Guinea language, experience difficulties in counting that are related to symmetries in the body part used to represent numbers. Languages such as Oksapimin can support counting to a limited degree but cannot support the more complicated concepts and operations of mathematics, such as division or the distinction between rational and irrational numbers.

Subtler but real differences in arithmetic capability are associated with linguistic differences in fully literate societies. English has a reasonably regular system for number naming, except that the numbers 11 and 12 are special and 13 to 19 are irregular compounds. The derivation of decade numbers is also somewhat obscure. Thus an English speaker must learn to manipulate 13 primitive terms (0–12), 7 "teen" terms with a special systematicity, 9 decade terms (20, 30, . . . 100), and a small number of special large-number terms, such as thousand, million, and billion. In contrast, the structure of Chinese terms for numbers maps exactly onto the structure of numbers expressed in modulo-ten arithmetic. A Chinese child only has to memorize 11 basic terms (for 0–10) and 3 special terms for 100, 1000, and 10,000. The number 11, in Chinese, is literally "ten and one," whereas an English speaker must use the special term "eleven." Italian is similar to the English system, except that there is less systematicity for names of the numbers 11–19. When English-speaking children learn arithmetic they have difficulty learning how to count in the teen range, but Chinese-speaking children do not (Miller & Stigler, 1987). Agnoli and Zhu (1989) found that Italian-speaking children show difficulties similar to those seen in English-speaking children.

Spatial schema provide another example of how linguistic differences in restricted schema may influence thought. Every human being lives in a three-dimensional space that, for virtually all everyday purposes, can be described by Euclidean geometry and the Cartesian notation. As English-speaking people become familiar with a new space their knowledge changes

from a procedural representation, in which they think about relative locations in terms of the movements required to get from one position to another, to a surveyor's representation, in which locations are thought of as points in a two- or three-dimensional Euclidean space. This progression is not a casual feat. Some individuals never develop a surveyor's representation of familiar environments (Thorndyke & Stasz, 1980).

Modern European languages support both procedural and survey representations. We can say right and left and we can also say north, east, south, and west. There are other languages that do not provide these distinctions. Mixtec (Lakoff, 1987) and Quechua (Bastien, 1978) are examples.

The Quechua are a particularly interesting case. They live in an exceptionally mountainous terrain in the Andes. There is no word for *flat* in Quechua, which must make thinking about a plain difficult. The Quechua give directions in terms of body images that emphasize verticality. For instance, they will speak of the head, belly, or foot of a mountain. The system can be complex, because villages are organized into a system of body-part relations, so that a given village may be the foot with respect to a village above it and the head with respect to a village below it. The Quechua language thus has the disadvantage of not being suitable for an unfamiliar terrain. Terms like *belly*, when used as metaphors for inanimate objects, rely on shared knowledge about the environment. Speakers of languages that use the cardinal points for direction can unambiguously designate the "north side" of a mountain they have never seen, but the "face" of a mountain is meaningless unless the communicators share a convention about the mountain being discussed.

At present, the evidence for our assertions is largely based on anthropological field reports and reasonable inferences. However, the issue is one that can be studied empirically. Do Quechua and Mixtec speakers develop surveyor representations? If not, can they be taught to do so without simultaneously teaching them a new language? Finally, are those information-processing characteristics that predict good spatial orientation in English speakers, and which imply an ability to manipulate Euclidean representations, equally predictive of the orientation ability of Quechua or Mixtec speakers?

Now let us look at some less restricted schema that deal with the world as it is believed to be rather than with the description of physical reality. Can language affect our ability to develop a representation of something that is not true? Bloom (1981) claimed that it can. He pointed out that the Chinese language does not have a subjunctive form and that Chinese speakers therefore have to resort to circumlocutions to express what is a straightforward English utterance. He offered the (apparently factual) example of a judge telling a visitor to the U.S., "If you weren't leaving tomorrow, you would be deportable." After some difficulty, the sentence was translated as "I know you are leaving tomorrow, but if you do not leave, you will be deported" (Bloom, 1981, p. 18). To an English speaker these two sentences are not equivalent.

Bloom reported studies that he interpreted as showing that Chinese speakers do poorly in formal studies of reasoning based on counterfactual statements. Au (1983) and Liu (1985) have questioned Bloom's conclusions on the grounds that the sentences used in Bloom's research were not well expressed in Chinese. Both Bloom and his critics speak as if the issue were whether or not a Chinese speaker can *understand* a counterfactual argument. We, here and elsewhere (Hunt & Banaji, 1988), argue that the issue is too sharply stated. The issue is not whether Chinese can perform counterfactual reasoning, but whether the relative cost of such reasoning is greater in Chinese than in English. We suggest that research that is based on reaction time or secondary task paradigms might shed more light on the controversy over Bloom's contention than can be shed by studies of error rates alone.

The use of the subjunctive is not the only example of a language that provides a linguistic schema for distinguishing between truth and speculation. In Quechua the speaker is required to make the distinction by using a different suffix to indicate accounts that are based on hearsay or that are directly acquired knowledge. In fact, there are several different validational suffixes that are intended to express how certain the speaker is about the communication (Jake & Chuquin, 1979). What this does is change a major problem-solving aspect of conversation from a listener's task to a speaker's task.

Now let us move to the social world. This is an area where, in a sense, the most interesting Whorfian effects should be found, but it is also the area where precise measurements are hardest to obtain.

Language provides the coding system for transmission of an idea from one person to another. The codes must refer to prototypes. Spontaneous labeling by language users will influence their memory for social or ill-structured perceptual events. C. Hoffman, Lau, and Johnson (1986) showed descriptions of individuals to bilingual English–Chinese speakers. These descriptions were chosen so that the character traits actually presented were part of stereotypes in either language, but the stereotypes that one might infer would depend on whether the individual was labeled with a Chinese or English stereotype. Subsequently, people were asked if particular behaviors not in the original description were likely to be characteristic of the target person. English speakers extrapolated traits associated with their stereotypes. Chinese speakers addressed in Chinese used Chinese stereotypes, but when addressed in English they used English stereotypes. Whorf anticipated C. Hoffman et al. (1986) by 30 years when he observed that English speakers will smoke near gasoline drums labeled "empty," even though the fumes in empty drums are highly explosive.

Hoffman et al.'s (1986) methods could be extended to investigate a number of other Whorfian effects. The area of social rank distinction is of particular interest, because emphasis on rank varies widely across languages. Modern English speakers use the all-purpose "you" for both lovers and door-to-door salesmen. Other languages maintain a strong distinction between pronouns for intimate and formal address. There is at least anecdotal evidence that these cues are used for cognitive appraisal of a social situation. Wierzbicka (1985b) reported that Polish–English bilinguals who have previously spoken to each other only in English feel embarrassment when they must address each other in Polish, because they are not sure what form of address they should use. Had they talked only in Polish the question would have been resolved at the appropriate time.

Linguistic support for social perception is even more developed in Japanese, which has a complex system of honorifics (Harada, 1976). Does the need to choose an honorific alter the

user's perception of a social situation in a way not apparent to an English speaker? Suppose that a person described a group of people to an English–Japanese bilingual and used English to describe some people and Japanese to describe others. The people described would differ both in the honorific used and other aspects (e.g., age, professional interests, and gender). Only some of these differences would be associated with the choice of honorific. Next, the bilingual would be asked to group the individuals by perceived similarities. Would multidimensional scaling reveal changes in the representational structure depending on the language used in the initial description? More crucially, could the differences in structure of the mental representation be related to the differences in the language's system of honorifics?

Systems of honorifics reflect the presence of schema within a culture, because the purpose of the honorific is to identify the filler for a particular slot within a schema. Pragmatic effects, which were little discussed in Whorf's time, provide another way of alerting a listener to the applicability of a culturally relevant schema. This is well illustrated by the use of indirect requests, as in "Can you open the door?" Such statements are normally requests for action rather than information. Wierzbicka (1985a, 1985c) argued that the form and prevalence of indirect requests in English are related to the personal autonomy schema important in British and American societies. Indirect requests are much less common when personal autonomy schema are less relevant. In particular, Wierzbicka claimed that Slavic societies emphasize solidarity and conformity to group mores more than individual autonomy. According to her, a polite Pole would tell a respected dinner guest to "Sit, sit!" for dinner, a form of address that English speakers reserve for dogs (Wierzbicka, 1985a, pp. 146–147). The same remark could be made of the Quechua, who do not value privacy to nearly the extent that Americans do. Instead of saying "Can you open the door?" in a modulated tone a Quechua would order (C. Greenway, personal communication, 1989).

We believe that these naturalistic observations make a fairly strong case for the role of pragmatics in shaping thought. They could be extended by experimental studies. Systematic observations could be made of the frequency of use of indirect forms in different social situations. For instance, within the English-speaking community does the frequency of use of direct or indirect requests vary depending on the relevance of the autonomy schema in a particular social situation? Suppose knowledgeable experts were asked to order societies by their emphasis on personal autonomy. Would this correlate with the use of indirect requests in those societies? We regard this as a particularly important study, because the present literature indicates that speech acts imply a value for personal autonomy. What we do not know is whether or not the use of indirect speech acts accentuates the salience of personal autonomy.

Conclusion

Arguments in support of the Whorfian hypothesis can degenerate into a potpourri of discussions of specific examples of linguistic influences on cognition. Whorf's own writings can be criticized on this ground. We have attempted to systematize the discussion by considering the Sapir–Whorf approach from the perspective of cognitive psychology. We certainly have not exhausted the topic. For instance, we have not considered crosscultural comparisons of language learning or the psychological consequences of bilingualism. There are substantial literatures on both topics. Delving into them would simply take us too far afield.

We believe that every utterance in language *A* has a translation in language *B*. This does not make the Sapir–Whorf hypothesis untrue. The issue is one of cost: Are there statements that are natural in language *A* that are stateable but unmanageable in language *B*? The Whorfian hypothesis is properly regarded as a psychological hypothesis about language performance and not as a linguistic hypothesis about language competence. Our review has convinced us that different languages pose different challenges for cognition and provide differential support to cognition.

We are aware that we have provided selected examples rather than the results of an extensive program of research. We have tried to be systematic in our selection, progressing from lexical effects through syntax and semantics to representational and pragmatic effects. The next step will be to develop empirical research in an even more systematic and extensive manner. That research will certainly have to include both controlled experimental studies and naturalistic observations of language use (i.e., a combination of experimental psychology, comparative linguistics, and anthropology). Language is too human to be confined to a single discipline.

References

Aaronson, D., & Ferres, S. (1986). Sentence processing in Chinese–American bilinguals. *Journal of Memory and Language, 25*, 136–162.

Agnoli, F., & Zhu, J. J. (1989, April). *One, due, san: Learning to count in English, Italian, and Chinese.* Paper presented at the meeting of the Society of Research in Child Development, Kansas City, MO.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Au, T. K. -F. (1983). Chinese and English counterfactuals: The Sapir–Whorf hypothesis revisited. *Cognition, 15*, 155–187.

Bastien, J. W. (1978). *Mountain of the condor.* St. Paul, MN: West.

Bates, E., & Devescovi, A. (1989). Crosslinguistic studies of sentence production. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 225–253). Cambridge, England: Cambridge University Press.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). Cambridge, England: Cambridge University Press.

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition, 11*, 245–299.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley, CA: University of California Press.

Bloom, A. (1981). *The linguistic shaping of thought: A study of the impact of language on thinking in china and the west.* Hillsdale, NJ: Erlbaum.

Bowerman, M. (1985). What shapes children's grammars? In D. I. Slo-

bin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues* (pp. 1257–1319). Hillsdale, NJ: Erlbaum.

Caelli, T. (1981). *Visual perception.* Elmsford, NY: Pergamon Press.

Carroll, J. B. (1963). Linguistic relativity, contrastive linguistics and language learning. *International Review of Applied Linguistics, 1,* 1–20.

Carroll, J. B., & Casagrande, J. B. (1958). The function of language classification in behavior. In E. E. Maccoby, T. Newcomb, & E. L. Hartley (Eds), *Readings in social psychology* (3rd ed., pp. 18–31). New York: Holt, Rinehart & Winston.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics.* San Diego, CA: Harcourt Brace Jovanovich.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30,* 73–105.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 561–584.

Einstein, A. (1954). *Ideas and opinions.* New York: Crown.

Ericsson, A. (1985). Memory skill. *Canadian Journal of Psychology, 39,* 188–231.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech–understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys, 12,* 213–253.

Fishman, J. A. (1980). The Whorfian hypothesis: Varieties of valuation, confirmation and disconfirmation: 1. *International Journal of the Sociology of Language, 26,* 25–40.

Foss, D. J., & Hakes, D. T. (1978). *Psycholinguistics. An introduction to the psychology of language.* Englewood Cliffs, NJ: Prentice-Hall.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 554–586). Hillsdale, NJ: Erlbaum.

Gass, S. M. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics, 8,* 329–350.

Greenberg, J. H. (1978). Generalizations about numeral systems. In J. H. Greenberg (Ed.), *Universals of human language: Vol. 3. Word structure* (pp. 249–295). Stanford, CA: Stanford University Press.

Harada, S. I. (1976). Honorifics. In M. Shibatani (Ed.), *Japanese generative grammar.* San Diego, CA: Academic Press.

Harrington, M. (1987). Processing transfer: Language-specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics, 8,* 351–377.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3,* 275–310.

Heider, E., & Olivier, D. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology, 3,* 337–354.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8,* 176–184.

Hoffman, C., Lau, I., & Johnson, D. R. (1986). The linguistic relativity of person cognition. *Journal of Personality and Social Psychology, 51,* 1097–1105.

Hoffman, E. (1989). *Lost in translation: A life in a new language.* New York: Dutton.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review, 85,* 109–130.

Hunt, E. (1980). Intelligence as an information processing concept. *Journal of British Psychology, 71,* 449–474.

Hunt, E., & Banaji, M. R. (1988). The Whorfian hypothesis revisited: A cognitive science view of linguistic and cultural effects on thought. In J. W. Berry, S. H. Irvine, & E. B. Hunt (Eds.), *Indigenous cognition: Functioning in cultural context* (pp. 57–84). Dordrecht, The Netherlands: Martinus Nijhoff.

Hunt, E., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7, pp. 87–121) San Diego, CA: Academic Press.

Hunt, E., & Lansman, M. (1986). Unified model of attention and problem solving. *Psychological Review, 93,* 446–461.

Hunt, E., & Poltrock, S. E. (1974). The mechanics of thought. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 277–350). Hillsdale, NJ: Erlbaum.

Jake, J. L., & Chuquin, C. (1979). Validation suffixes in Imbabura Quechua. *Papers from the Regional Meetings, Chicago Linguistic Society, 15,* 172–184.

Just, M., & Carpenter, P. (1987). *The psychology of reading and language comprehension.* Needham Heights, MA: Allyn & Bacon.

Kail, M. (1989). Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence comprehension* (pp. 77–117). Cambridge, England: Cambridge University Press.

Kameyama, M. (1987). *Computing Japanese discourse: Grammatical disambiguation with centering constraints.* Paper presented at the workshop on Computing Japanese at the University of Manchester Institute of Science and Technology, Manchester, England.

Kay, P., & Kempton, W. (1984). What is the Sapir–Whorf hypothesis? *American Anthropologist, 86,* 65–79.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech, 32,* 1–23.

Kilborn, K., & Ito, T. (1989). Sentence processing strategies in adult bilinguals. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 257–291). Cambridge, England: Cambridge University Press.

Klahr, D., Langley, P., & Neches, R. (1987). *Production systems models of learning and development.* Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lantz, D., & Stefflre, V. (1964). Language and cognition revisited. *Journal of Abnormal and Social Psychology, 69,* 472–481.

Liu, L. G. (1985). Reasoning counterfactually in Chinese: Are there any obstacles? *Cognition, 21,* 239–270.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement scaling and psychophysics. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., pp. 3–74). New York: Wiley.

Lucy, J. A., & Shweder, R. A. (1979). Whorf and his critics: Linguistic and nonlinguistic influences on color memory. *American Anthropologist, 81,* 581–615.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing.* Cambridge, England: Cambridge University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23,* 127–150.

MacWhinney, B., Pléh, Cs., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology, 17,* 178–209.

McDonald, J. L. (1987a). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics, 8,* 379–413.

McDonald, J. L. (1987b). Assigning linguistic roles: The influence of conflicting cues. *Journal of Memory and Language, 26,* 100–117.

Miao, X. (1981). Word order and semantic strategies in Chinese sentence comprehension. *International Journal of Psycholinguistics, 8,* 23–33.

Miller, K. F., & Stigler, J. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Development, 2,* 279–305.

Mitchell, D. C., Cuetos, F., & Zagar, D. (1990). Reading in different languages: Is there a universal mechanism for parsing sentences? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 285–302). Hillsdale, NJ: Erlbaum.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience. The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18,* 331–340.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Il nuovo ragazzini [dictionary]. (1984). Bologna, Italy: Nicola Zanichelli, S. P. A.

Perfetti, C. A., Beverly, S., Bell, L., Rodger, K., & Faux, R. (1987). Comprehending newspaper headlines. *Journal of Memory and Cognition, 26,* 692–713.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 58,* 193–198.

Pinker, S. (1990). Language acquisition. In D. N. Osherson & H. Lasnik (Eds.), *Language: An invitation to cognitive science* (Vol. 1, pp. 199–241). Cambridge, England: Cambridge University Press.

Pléh, C. (1989). The development of sentence interpretation in Hungarian. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 158–184). Cambridge, England: Cambridge University Press.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Rheingold, H. (1988). *They have a word for it.* Los Angeles: Jeremy P. Tarcher.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1–20.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4,* 328–350.

Rosch, E. (1974). Linguistic relativity. In A. Silverstein (Ed.), *Human communication: Theoretical perspectives* (pp. 254–279). New York: Halsted Press.

Sapir, E. (1941). *Language, culture, and personality: Essays in memory of Edward Sapir* (L. Spier, Ed.). Menasha, WI: Sapir Memorial Publication Fund.

Saxe, G. B., & Posner, J. K. (1983). The development of numerical cognition: Cross-cultural perspectives. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 291–317). San Diego, CA: Academic Press.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36–71.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27,* 379–423.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues* (pp. 1157–1256). Hillsdale, NJ: Erlbaum.

Slobin, D. I. (1989, July). *Factors of language typology in the crosslinguistic study of acquisition.* Paper presented at the Tenth Biennial Meetings of the International Society for the Study of Behavioral Development, Jyväskylä, Finland.

Slobin, D. I., & Bocaz, A. (1988). Learning to talk about movement through time and space: The development of narrative abilities in Spanish and English. *Lenguas Modernas, 15,* 5–24.

Sokolov, J. L. (1989). The development of role assignment in Hebrew. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 118–157). Cambridge, England: Cambridge University Press.

Sperber, P., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford, England: Basil Blackwell.

Stefflre, V., Castillo Vales, V., & Morley, L. (1966). Language and cognition in Yucatan: A cross cultural replication. *Journal of Personality and Social Psychology, 4,* 112–115.

Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology, 12,* 137–176.

Twain, M. (1977). The awful German language. In C. Neider (Ed.), *A tramp abroad* (pp. 63–81). New York: Harper & Row. (Original work published 1880).

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* San Diego, CA: Academic Press.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf.* New York: Wiley.

Wickens, D. D. (1972). Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 191–216). Washington, DC: Winston & Sons, Inc.

Wierzbicka, A. (1985a). Different cultures, different languages, different speech acts. *Journal of Pragmatics, 9,* 145–178.

Wierzbicka, A. (1985b). The double life of a bilingual. In R. Sussex & J. Zubrzycki (Eds.), *Polish people and culture in Australia* (pp. 187–223). Canberra, Australia: Australian National University.

Wierzbicka, A. (1985c). A semantic metalanguage for a crosscultural comparison of speech acts and speech genres. *Language in Society, 14,* 491–514.

Wulfeck, B., Juarez, L., Bates, E., & Kilborn, K. (1986). Sentence interpretation in healthy and aphasic bilingual adults. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives* (pp. 199–219). Hillsdale, NJ: Erlbaum.

Yee, P. L., Hunt, E., & Pellegrino, J. W. (in press). Co-ordinating cognitive processing: Individual differences and a general model. *Cognitive Psychology.*

Received January 10, 1990
Revision received November 6, 1990
Accepted December 13, 1990 ■