CHAPTER 38

COGNITIVE LINGUISTICS AND LINGUISTIC RELATIVITY

ERIC PEDERSON

1. Introduction

Linguistic relativity (also known as *the Sapir-Whorf Hypothesis*) is a general cover term for the conjunction of two basic notions. The first notion is that languages are *relative*, that is, that they vary in their expression of concepts in noteworthy ways. What constitutes "noteworthy" is, of course, a matter of some interpretation. Cognitive scientists interested in human universals will often describe some particular linguistic variation as essentially minor, while others, for example, some anthropological linguists, may describe the same variation as significant.

The second component notion to linguistic relativity is that the linguistic expression of concepts has some degree of influence over conceptualization in cognitive domains, which need not necessarily be linguistically mediated. In textbooks, this notion of language affecting conceptualization is typically divided into "strong" and "weak" hypotheses. The "strong" hypothesis (also known as *linguistic determinism*) is that the variable categories of language essentially control the available categories of general cognition. As thus stated, this "strong" hypothesis is typically dismissed as untenable. The "weak" hypothesis states that the linguistic categories

may influence the categories of thought but are not fundamentally restrictive. As thus stated, this "weak" hypothesis is typically considered trivially true.

Arguably, this simplification of the broad issue of the relationship between linguistic and cognitive categorization into two simple ("strong" vs. "weak") statements has impeded development of genuinely testable hypotheses and has helped lead studies of linguistic relativity into academic ill-repute. Modern research into the general question of linguistic relativity has focused on more narrowly stated hypotheses for testing, that is, investigating the specific relationships between particular linguistic categories (e.g., the categories of number, color, or spatial direction) and more exactly specified cognitive operations (e.g., encoding into long-term memory or deductive reasoning).

This chapter is organized as (i) a brief history of the research question (section 2); (ii) a discussion of the challenges in designing research into linguistic relativity (section 3); (iii) the treatment of linguistic relativity within works generally representative of Cognitive Linguistics (section 4); and (iv) a survey of classic and more modern (pre- and post-1980s) research within linguistics, anthropology, and psychology (section 5).

In addition to this chapter, several other surveys of linguistic relativity may be consulted. Lucy (1997a) gives a broad overview of different approaches which have investigated linguistic relativity, while Lucy (1992b) elaborates on a particular empirical approach and provides detailed critiques of previous empirical work. Lee (1996) provides historical documentation to the often poorly understood work of Benjamin Lee Whorf (see also Lee 2000). Hill and Mannheim (1992) trace the history of the notion of world view with respect to language through twentieth-century anthropology, from Boas through Cognitive Linguistics of the 1980s to the work of John Lucy. Hill and Mannheim also provides a useful overview of the anthropological cum semiotic approach to culturally embedded language use—see especially Hanks (1990) and Silverstein (1985, 1987).

Smith (1996) also discusses the writings of Sapir and Whorf to clarify that most popular accounts of the Sapir-Whorf Hypothesis are not directly derivative of their work. She is also concerned that the relatively large-scale dismissal of the Sapir-Whorf Hypothesis in academic culture has been at the expense of serious research into the relationships between language and thought. Similar discussion of the "demise" of the "Whorf Hypothesis" and the misconstrual of Whorf's actual writings can be found in Alford (1978).[1]

Koerner (2000) also provides a survey of the "pedigree" of linguistic relativity "from Locke to Lucy," that is, from the seventeenth through the twentieth century. Chapters 10–12 of Foley (1997) as well provide historical coverage of the notion, with summaries of fairly recent work with spatial language and classifiers. Duranti (1997) similarly provides historical coverage with particular emphasis on the American anthropology traditions.

Hunt and Agnoli (1991) revisit linguistic relativity from the perspective of cognitive psychology, which had largely rejected the notion as either

false or uninteresting during the 1970s. Within canonical Cognitive Linguistics, Lakoff (1987) dedicates chapter 18 of *Women, Fire, and Dangerous Things* to discussions of evidence for and types of linguistic relativity. Many of the principles from that chapter have informed the remainder of his work.

2. Historical Speculation and Modern Formulations

Given the wealth of historical surveys of linguistic relativity, this chapter will focus more on modern work and methodological issues. However, a brief overview of the history of linguistic relativity theorizing will help to situate the modern research questions.

2.1. From Humboldt through Whorf

The most widely cited intellectual antecedent for linguistic relativity is the work of Humboldt. Later, the work of Boas is widely seen as the inheritor of the Humboldtian notions and through him, the concern with linguistic relativity was taken up in the writings of Sapir, who developed the vital notion of the "patterns" or structural systematicity of language as being particularly relevant to the relationship between language, mind, and culture.

Humboldt's principal work addressing linguistic relativity is *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlecht* [On the diversity of human language construction and its influence on the mental development of the human species]. There are many editions and translations of this work; for a recent edition of Peter Heath's English translation, see Losonsky (1999). The philosophical precursors to Humboldt, as well as linguistic relativity in general, is discussed in Manchester (1985), and an overview of Humboldt's notion of language and *Weltansicht* ('world view') is provided in Brown (1967).

The writings of Benjamin Lee Whorf are best known through Carroll's edited collection Whorf (1956). This collection helped to popularize the notion that the categories of language may influence the categories of thought. However, Lee (1996) argues—especially in light of the previously unpublished "Yale report" (see Whorf and Trager [1938] 1996)—that Whorf was concerned with the interpenetration of language and thought; that is, the two words *language* and *thought* refer to aspects of a single system, and it is a misapprehension to ask in what way one affects the other. This is quite distinct from the more modular view of language processing dominant in current psychology and linguistics.

2.2. Literacy

While modern linguistics places considerable emphasis on *spoken* language—which means that this chapter will focus on the potential cognitive impact of the categories found in spoken or signed languages—the role of *literacy* to cognitive and cultural development has long been a subject of debate.

Early twentieth-century experiments on the relationship between literacy and cognitive development were conducted by Aleksandr Luria and colleagues (for an overview in English, see Luria 1976). This classic work investigated the effects of previously established, Soviet-era adult literacy programs on the development of various cognitive skills. There were a number of methodological problems with that work—perhaps the most significant one being the confounding of formal schooling with the acquisition of literacy (or conversely, the lack of formal schooling with nonliterate populations). The largest single effort to overcome this common confound is reported by Scribner and Cole (1981), who investigated effects of literacy acquisition in the absence of formal schooling. The designs and subject pools were still not completely free of confounding factors and the results, while fascinating, give a largely mixed picture of the effects of literacy as an independent factor on cognition.

"The literacy hypothesis," namely that various cultural features can be traced to the development of literacy in the history of a given culture, has been subject to considerable debate. Goody and Watt (1962), one of the better known works, extolled the effects of specifically *alphabetic* literacy as critical in the development of early Greek and later European culture. This view came under considerable criticism, and Goody himself later backed away from the specific claims about alphabetic literacy.[2] However, on a more general level, the claim that literacy engenders certain cognitive changes—especially enhanced metalinguistic awareness—continues to be argued. Readers interested in the effects of literacy on cognition could also consult Scinto (1986), Graff (1987), Olson (1991, 2002), Ong (1992), and references therein.

Rather than studying the general effects of reading and writing on cognition, one line of research has been concerned with the effects of learning particular writing systems. Morais et al. (1979) investigate the effects of child-acquired literacy on phonemic awareness, and Read et al. (1986) present evidence arguing that alphabetic literacy, but not logographic and syllabic literacy, leads to phonemic awareness. In Danziger and Pederson (1998) and Pederson (2003), I argue that familiarity with specific graphemic qualities can lead to differences in visual categorization in nonwriting/nonreading tasks.

2.3. Folk Classification

Anthropologists have long been concerned with *folk classification*, that is, the culturally specific ways in which linguistic and other categories are organized into coherent systems. Perhaps the richest body of work is in the area of taxonomies of

natural kinds (plants, animals, etc.). This research is conveniently served by having a scientific standard for comparison. While there is abundant anecdotal evidence that people interact with natural kinds according to their taxonomical relations to other natural kinds (e.g., *X* is a pet, so treat it like other pets), there has not been much in the way of psychological-style testing of specific linguistic relativity hypotheses in this domain. For an introduction to folk classification, see Hunn (1977, 1982), Berlin, Breedlove, and Raven (1973), Berlin (1978), and Blount (1993).

2.4. Formulations of Linguistic Relativity

There are many semantic domains one could search for linguistic relativity effects—that is, domains in which one might find linguistic categories conditioning nonlinguistic categorization. For example, cultures and languages are notorious for having varying kinship terms, which group into major types with various subtypes. Importantly, the categories of allowable behaviors with kin tend to correspond to the grouping by kinship terminology. For example, South Indian (Dravidian) languages systematically distinguish between cross-cousins and parallel cousins, with marriage allowed between cross-cousins and incest taboo applying to parallel cousins. In contrast, North Indian languages typically classify all cousins with siblings and incest taboo applies to all (see Carter 1973).

However important sexual reproduction may be to our species, the standards of marriage are clearly the result of cultural convention overlaid on biological predispositions. Accordingly, finding linguistic variation corresponding to categories of human behavior in such a domain is not generally taken as a particularly revealing demonstration of linguistic relativity. Likewise, elaborated vocabulary sets in expert domains and impoverished sets where there is little experience, however interesting, are also not taken as particularly revealing. While a tropical language speaker may lack the broad vocabulary of English for discussing frozen precipitation, that same speaker may be quite particular in distinguishing what English speakers lump together as 'cousins'.

In other words, cases of categorization which are dependent on environmentally or culturally variable experience are generally considered uninteresting domains for the study of linguistic relativity. This corresponds to the late twentieth-century bias toward universalism in the cognitive sciences; namely, for variation to be noteworthy, it should be in a domain where variation was not previously thought to be possible. That is to say, for linguistic relativity to be broadly interesting, it must apply within cognitive domains which operate on "basic" and universal human experience.

3. Challenges in Researching Linguistic Relativity

3.1. Intralinguistic Variation

Speakers may use language differently across different contexts, and this difference may be indicative of shifting conceptual representations. One of the few studies within Cognitive Linguistics to empirically address intralinguistic variation is Geeraerts, Grondelaers, and Bakema (1994, especially chapter 4: "Onomasiological Variation"), which explores alternative expressions as the representation of different construals and perspectivization.

Of course, some of these alternative expressions may be confined to some subcommunities and dialects. While linguistic relativity is typically discussed as the difference across speakers of distinct languages, there is every reason to wonder about parallels with differences in conceptualization that may exist within a single language community. Speakers of different dialects may have different linguistic patterns which might be hypothesized to correspond to different habitual conceptualizations. In Pederson (1993, 1995), I investigate communities of Tamil speakers who systematically vary in their preference for terms of spatial reference, but who otherwise speak essentially the same dialect.

The work of Loftus (1975) has demonstrated that the choice of particular linguistic expressions at the time of encoding or recall may well influence nonlinguistic representation of events. Extrapolating from Loftus's work, we might wonder to what extent language generally can prime specific nonlinguistic representations—I call this the *language as prime* model. The fact that social humans are surrounded by linguistic input suggests that there might be a cumulative effect of this language priming. Indeed, if a particular linguistic encoding presented before a certain perception influences the nonlinguistic encoding or recall of that perception, what then might be the cumulative effect of one type of linguistic encoding rather than another being used throughout a speaker's personal history? If, for example, the classifiers of a speaker's habitual language force categorization of certain objects as 'long and thin', it seems reasonable that such objects may be remembered as potentially longer or thinner than they actually were.

Of course, if there were no consistent pattern to the linguistic priming, then we would not expect any single representation to become dominant. Indeed, Kay (1996) has argued that there is considerable flexibility within any language for alternative representations, and speakers may well alternate from one representation to another. This suggests that rather than a single and simple "world-view" necessary for a cleanly testable hypothesis, speakers may draw on complex "repertoires" of representations. While this does not preclude the possibility of systematic differences across languages having different repertoires, it certainly argues that the differences are far less obvious.

Given flexibility within a single language, a linguistic relativity hypothesis to be tested may need to compare patterns which are pervasive in one language and underexpressed in another language. This can be difficult to compensate for in an experimental design. A balanced design might seek opposing, but functionally equivalent systems, which are dominant in each language community. Each community may have both systems in common, but not to the same level of default familiarity. Of course, the experimental measure needs to be sufficiently nonpriming itself so as to allow each subject population to rely on their default mode of representation.

3.2. Selecting a Domain

Universals in categorization may be of more than one type. Most relevantly, some categories may be essentially innate, that is, an internal predisposition of the organism. Other universal categories may be the result of commonalities of all human environments in conjunction with our innately driven mechanisms. Even assuming that we can reliably presume that certain categories are universal, determining which are purely innate and which derive from interaction with universal properties of the environment is not a trivial task.

Variation in innate properties is impossible—except inasmuch as the variation is within innately proscribed limits—so we cannot look for linguistic relativity effects in these domains. For linguistic relativity effects to be both interesting to cognitive scientists and robust in their operations, they must apply in a domain which is generally presumed universal by virtue of the common environment, but which can be hypothesized to be nonuniversal. As discussed above, demonstrating effects from language type in cognitive domains with wide variation is unexciting. It follows that the researcher interested in testing linguistic relativity best seeks a domain which is hypothesized to be fairly basic to cognition, but just shy of exhibiting a universal pattern.

This motivates modern linguistic relativity studies to examine categorization in domains presumed to derive somewhat immediately from basic perceptual stimuli or fundamental mechanisms of reasoning. The majority of such empirical studies concern categorization of visual or spatial properties of objects or the environment. A few studies have examined purported differences in reasoning, but these are inherently more difficult to pursue. Object properties and the environment can be experimentally controlled, but processes of reasoning—especially in cross-cultural work—are notoriously difficult to measure while maintaining adequate control of subject variables.

3.3. Independent Evidence for Language and Cognition

Linguists—especially cognitive linguists—frequently claim that a particular linguistic form represents a particular underlying conceptualization. Obviously, however, any substantial claim of a relationship between language and cognition needs independent assessment of each and a correlation established between the two.

Perhaps surprisingly, most work on linguistic relativity spends remarkably little effort demonstrating the linguistic facts prior to seeking the hypothesized cognitive variable. Some of the most severe criticisms of linguistic relativity studies have worried about this insufficient linguistic description. Lucy (1992b) is especially clear in his call for more careful linguistic analysis preparatory to linguistic relativity experimentation.

Given the relative accessibility of the linguistic facts compared with the difficulty inferring cognitive behavior from behavioral measures, one could argue that the often minimal characterization of language is of unacceptable sloppiness. More charitably, linguistic facts are typically quite complex, and in an effort to seek a testable hypothesis, a certain amount of simplification becomes inevitable. Unfortunately, there is no standard to use in evaluating the adequacy of a linguistic description for linguistic relativity work other than using the general standards of descriptive linguistics. Descriptive linguistics tends to be as exhaustive as is practically possible and does not necessarily foster the creation of simple hypotheses about linguistic and conceptual categorization. On the other hand, it is difficult to argue that studies in linguistic relativity should hold their linguistic descriptions to a lower standard.

A related problem is the variability of language. Since many different varieties of language exist depending on communicative and descriptive context, it can be quite misleading to speak of Hopi or English as having a specific characteristic, unless one can argue that this characteristic is true and uniquely true (e.g., there are no competitive constructions) in all contexts. This is, needless to say, a difficult endeavor, but failing to argue the general applicability of the pattern invites the next linguist with expertise in the language to pull forth numerous counterexamples. Studies most closely following the approaches advocated by Whorf have tended to focus on basic grammatical features of the language which are presumed to be fairly context independent. However, this may overlook other linguistic features which may well be relevant to a particular hypothesis of linguistic and conceptual categorization.

One way to partially circumvent this problem was followed in Pederson et al. (1998), which seeks to describe language characteristics typically used for, in this case, table-top spatial reference. There is no attempt to include or exclude information on the basis of whether or not the relevant language elements were grammaticized or lexicalized. Rather, if the information was present in the language used for a particular context, these linguistic categories are presumed to be available conceptual categories within same or similar contexts. This approach leaves unanswered the question of how broadly the linguistic description (or for that matter the cognitive description as well) applies to the subject population in a variety of

other contexts, but it does help ensure that the linguistic description is the most exact match for the cognitive enquiry.

3.4. Subvocalization or What Is Nonlinguistic?

If independent measures are to be taken of both language use and cognitive processes, then great care is necessary to ensure that the behavioral measure for the nonlinguistic cognitive process is not covertly measuring linguistically mediated behavior.

Ideally, the entire cognitive task would be nonlinguistic, but as a practical minimum, the instructions and training for the task must be couched in language which is neutral with respect to the current hypothesis. This is particularly difficult to manage when a language has grammatically obligatory encoding. How do we interpret an effect which may be due to obligatory encoding in the instructions? Is this just an effect of the instructions, or can we interpret this as a general language effect because the instructions only exemplify the continual linguistic context the subjects live within?

There is a general presumption that instructions to the subjects should be in the subjects' native language. One might be tempted to use a shared second language as a type of neutral metalanguage for task instructions, but this introduces unexplored variables. If there is the possibility of a cognitive effect from the regular use of one's native language, then there is also the possibility of an effect from the immediate use of the language of instruction. Additionally, it is more difficult to be certain that all subjects understand the second-language instructions in exactly the same way as the experimenter. Finally, it is unclear how one would guarantee that the language of instruction is neutral with respect to anticipated behavioral outcomes. The very fact that it may mark different categories from the native language may influence the outcome in unpredictable ways.

It is safest therefore to minimize any language-based instruction. General instructions (e.g., "Sit here") cannot be excluded, but critical information is best presented through neutral examples with minimal accompanying language. Since a dearth of talking makes it more difficult to monitor subject comprehension, it is imperative that the experimental design include a built-in check (e.g., *control trials*) to ensure that each subject understands the task in the same way—except, of course, for the variation for which the task was designed to test. An account of the effects of subtle changes in instruction with children in explorations with base ten number systems can be found in Saxton and Towse (1998).

Another concern is that subjects involved in an ostensibly nonlanguage measure actually choose to use language as part of the means of determining their behavior. For example, the subjects may subvocalize their reasoning in a complex problem and then any patterning of behavior along the lines of the linguistic categories is scarcely surprising. In Pederson (1995), I address this concern by arguing that if subjects have distinct levels of linguistic and conceptual representa-

tions, they should only choose to approach a nonlinguistic task using linguistic means if there were a sufficiently close match between these two levels with respect to the experiment. In effect, a subject's unforced decision to rely on linguistic categories can be understood as validation of at least one sort of linguistic relativity hypothesis.

3.5. Finding Behavioral Consequences of Linguistically Determined Cognitive Variation

Variation in categorization of spatial or perceptual features can be of relatively minor consequence. Whether one thinks of pencils more fundamentally as tools or as long skinny objects has probably little effect on their employment.

The most basic features of humans and their environment are stable across linguistic communities. Gravity pulls in a constant direction, visual perception is roughly comparable, and so forth. If there are cognitive differences across communities with respect to universal features, then these different cognitive patterns must have *functional equivalence*; that is, different ways of thinking about the same thing must largely allow the same behavioral responses. For example, whether a line of objects is understood as proceeding from left to right or from north to south makes little difference under most circumstances. If the objects are removed and the subject must rebuild them, either understanding of the array will give the same rebuilt line with no effect on accuracy. Accordingly, any experimental task must select an uncommon condition where the principle of functional equivalence fails to hold (see especially Levinson 1996). To continue this example, if the subject is rotated by 90 or 180 degrees before being asked to rebuild an array, the underlying representation (left-right or north-south) should result in a different direction for the rebuilding.

Without a context which effectively disambiguates the possible underlying representations from behavioral responses, a researcher must demonstrate that one subject population has a deficient or improved performance on a task and that this differential performance corresponds to a difference in (default) linguistic encoding. There is a long and sordid history of attributing deficiencies to populations that the investigator does not belong to. Accordingly, it is entirely appropriate that the burden of proof fall particularly hard on the researcher claiming that a studied population is somehow impaired on a given task as a result of their pattern of linguistic encoding. Even if the population is claimed to have an ability which is *augmented* by linguistic encoding, it is difficult to demonstrate that any difference in ability derives specifically from linguistic differences and not from any of a myriad of environmental (perhaps even nutritional) conditions.

Related to this is the concern for the *ecological validity* of the experimental task. A task may fail to measure subject ability or preferences owing to unfamiliarity of the materials, instructions, or testing context. Further, it is difficult to decide on the basis of just a few experiments which effects can be generalized to hold for nonexperimental contexts—to wit, the complexity of daily life. This is not,

however, an argument against experimentation as the inherently interpretive nature of simple observational data ultimately requires experimentally controlled measures.

3.6. Types of Experimental Design

Various types of experimental tasks have been used for investigating the cognitive side of linguistic relativity. Whatever research methods are used, reliability of the results is far more likely if there is triangulation from a number of observational and experimental methods.

Sorting and Triads Tasks

Perhaps the most common design used in linguistic relativity studies is a sorting task. Quite simply, the subject is presented with a number of stimuli and is asked to group them into categories. These categories may be ad hoc (subject determined) or preselected (researcher determined). Multiple strategies may be used for the sorting task, giving different sorting results. The most common variant of the sorting task is the triads task which presents a single stimulus to the subjects and asks them to group it with either of two other stimuli or stimuli sets; that is, does stimulus X group better with A or with B? (hence, the term *AXB test* in some research paradigms). For an archetypal example of a triads task, see Davies et al. (1998).

This task is easy to administer as long as the stimuli are reasonably tangible, interpretable, and able to be considered in a nearly simultaneous manner. One consideration of sorting designs is that subjects often report awareness of multiple strategies which might be employed. Of course, the researcher cannot indicate which is a preferred strategy and can only instruct the subject to sort according to "first impression," "whatever seems most natural," or other such instructions. The interpretation of these instructions may add an uncontrolled variable. Further, sorting tasks inherently invite the subjects to respond according to their beliefs about the researcher's expectations, which may not in fact be what would be the normal sorting decision outside of this task.

Discrimination Tasks

Other tasks seek to find different discriminations across populations. As a practical consequence, differences usually boil down to one population making finer or more distinctions than another population; see, for example, much of the work on color discrimination and linguistic labeling discussed in the debates in Hardin and Maffi (1997). However, it is at least theoretically possible that one population might be more sensitive to certain features at the expense of other features and that a contrasting population would show the reverse pattern.

A limitation of discrimination tasks is that for them to be interpretable, one must be able to assume that beyond the independent variable of different linguistic

systems, all subjects brought the same degree of attention, general task satisfying abilities, and so on to the experimental task. Should, for example, one population be less likely to be attentionally engaged, then this reduces the possibility of isolating a linguistic effect on cognition.

Problem Solving Tasks

Problem solving tasks are readily used in many types of research. In linguistic relativity studies, they are typically of two design types: difficult solution or alternative solution.

The first type involves a task which provides some difficulty in finding the solution. Some subjects are anticipated to be better or worse than others at solving the task. As with reduced discrimination just discussed, it is extremely difficult to argue that it is specifically the categories of language which lead to differential performance. The counterfactual reasoning task employed by Bloom (1981) was such a task, and the difficulty in interpreting its results was part of much of the controversy surrounding that work.

The second type of problem solving tasks allow for alternative solutions each of which should be indicative of a different underlying representation. As such, these are similar to triads tasks in that they allow each subject to find the most "natural" solution for them (at least within the given experimental context). For example, in Pederson (1995) I describe a transitivity task in which subjects know how each of two objects are spatially related to a third object. They must then decide which side of the second object the first/test object must be placed. Depending on how these relationships are encoded, the test object will be placed on a different side of the second object. Like triads tasks, there is the potential problem that the subjects may be aware of the possibility of multiple solutions, prompting responses derived from any number of uncontrolled factors.

Embedded Tasks

Within psychological research, there is a common solution to the problem of subject awareness of multiple possible responses. Namely, the actual measure of the task is *embedded* within another task for which the subject is more consciously aware. For example, subjects may be asked to respond as to whether a figure is masculine or feminine, but the researcher is really measuring the distribution of attention to the figures. While the embedded task may still be influenced by subject expectations, it is an indirect and presumably nonreflected influence. As such, one can argue that the responses measured by the embedded task are more likely to correspond to default behaviors used outside of this exact experimental context. The "Animals in a Row" task discussed in Pederson et al. (1998) was one such task, where subjects understood the task as one to recreate a *sequence* of toy animals, but the critical dependent measure was the *direction* the animals were facing when subjects placed them on the tabletop before them.

Variable Responses

The researcher must also be careful in coding fixed response types from the subjects. It may be that subject preference is for a response type not allowed by the forced choice, and when pigeonholed into a different response type, subjects may not be responding in a manner reflecting their typical underlying representations. Also, certain patterns (or lack of patterns) of responses may actually indicate a preference for a response type not anticipated by the experimental design. For example, in the "Animals in a Row" task just discussed, some populations—and not others—appear on the scoring sheets as preserving the orientation of the original stimuli roughly half the time. On closer inspection, many of these subjects were actually entirely consistent in giving the animals the same orientation (e.g., always facing left) regardless of the original orientation of the stimuli. Since the task appeared to be about the order and not the orientation of the animals, this is a perfectly reasonable response. Unfortunately, there was no hypothesis anticipating this response, and no claims could be made as to why some subjects and not others gave this response pattern.

3.7. Controlling Extraneous Variables

Work such as Kay and Kempton (1984) demonstrates that the effects of native language on nonlinguistic categorization tasks can vary with even slightly varied task demands. This is commonly interpreted as an indication that "relativity effects" are "weak." A more conservative interpretation is that there are many factors (of undetermined "strength") which can effect results and that language may be only one of many possible factors. The exact total effect of language will depend on what other nonlinguistic factors are in effect. This requires that an experimental design for linguistic relativity effects carefully control all foreseeable linguistic and nonlinguistic variables.

Linguistic Variables

Since they are most directly related to the tested hypothesis, language variables are perhaps the most critical to control in one's design.

Of fundamental importance is that one must be certain that the base language of the subjects is consistent with respect to whatever features have led to the specific hypothesis. This may seem trivial, but dialectal (and even idiolectal) variation may well have the effect that some speakers do not share certain critical linguistic features even though they ostensibly speak the same language.

Perhaps even more problematic is the issue of bilingualism. Unless all subjects are totally monolingual, this is a potential problem for the design. Generally, linguistic relativity tests presume that one's "native" language capacity is the most relevant, but this cannot preclude effects from other known languages. Age of acquisition of second languages may also vary widely; there is certainly no established model of the effects of age of acquisition on nonlinguistic category formation.

If nonnative categories have been learned, how can we assume that they are not also brought to bear on the experimental task—clouding the results in unpredictable ways? This is perhaps most insidious when the language of instruction differs from the native language. Suitably, then, serious work in linguistic relativity needs to use the native language for instruction, but even this is not necessarily a straightforward task. For example, how does one ensure that instructions to multiple populations are both exactly and suitably translated?

How to Control for Exact Translations in a Comparative Work?

Work in linguistic relativity has had an impact in translation theory. Indeed, belief in a sufficiently strong model of insurmountable language differences would suggest that complete translations would be difficult to attain. House (2000) presents an overview of the challenges of translation and suggests a solution to the problem of linguistic relativity and translation. Chafe (2000) also discusses translation issues with respect to linguistic relativity, and Slobin (1991, 1996) uses translations in his discussions of how languages most suitably express motion events (see the section on space, below). The work of Bloom and his critics (see the discussion below) is particularly relevant for this issue because the ability to translate the experimental task from English to Chinese was central to his research question of counterfactual reasoning. Indeed, one might be skeptical of any attempt to investigate linguistic relativity in which the nonlinguistic experimental design is essentially a language-based task.

Of immediate practical concern is the translation of instructions for any research instrument itself. It is difficult enough to be confident that two subjects speaking the same language have the same understanding of a task's instructions. How, then, can the researcher be confident that translations of instructions are understood identically by speakers of different languages especially in the context of an experiment which seeks to confirm that speakers of these different languages in fact do understand the world in different ways?

The most obvious solution is to avoid linguistic instruction entirely. This does not remove the possibility that subjects understand the task differently, but it does ensure that any different understanding is not the direct result of immediate linguistic context. However, there are severe restrictions on what can be reliably and efficiently instructed without language. Understandably, then, most research relies on language-based instruction. In such cases, one must seek to phrase instructions in such a way that one sample is not more influenced by the particular choice of phrasing than the other sample.

To invent an example, imagine we are interested in the effect of evidential marking (linguistic markings which indicate how information is known to the speaker) on the salience of sources of even nonlinguistic information to speakers of a language which obligatorily marks evidentiality. This population would contrast with speakers of a language which essentially lacks routine marking. How, then, might we word our instructions? Do we use expressions typical for each language

such that one set of instructions contains evidential marking and the other not? Alternatively, do we provide evidential information for both languages? In the case of the language which does not typically mark evidentials, providing this information would obviously be more "marked" in usage than for the other language. This greater markedness of the information might make the evidential information *more* salient for those subjects who normally do not concern themselves with any language expression of evidentiality, which in turn could make issues of evidentiality more salient than they would be under average conditions—countering the entire design of the experiment!

Recent Language Use

Another potential language factor affecting results might be preexperimental, but recent, language use. If the language of instruction can influence results, could not language use immediately prior to instruction also influence the results? Indeed, if we assume that linguistic categories prime access to parallel nonlinguistic categories, then how do we control for language use outside of the experimental setting? On the one hand, one could argue that language use outside of the experiment is exactly the independent variable under consideration, and this is controlled simply through subject selection. On the other hand, if a language has multiple ways of representing categories, what is the potential effect if a subject has most recently been using one of the less typical linguistic categories for his or her language? Once again, the cleanest solution to this risk is to test categories for which there is minimal linguistic variation within each of the examined languages.[3]

Conversation during Task

The last of the language variables to consider is language use during the experiment itself. Lucy and Shwedder (1988) found that forbidding subjects to have conversations between exposure and recall in a memory task allowed a greater recall of focal color terms than of nonfocal color terms (see the subsection on color below). Subjects who had (unrelated) conversations remembered focal and nonfocal colors about equally well. While Lucy and Shwedder do not provide a model for why this might be the case, it clearly suggests that even incidental language use during and perhaps around a task can have significant influences on performance. Other work (see Gennari et al. 2002) has suggested that even in cases where there might normally be no particular relation between habitual language use and performance on a nonlinguistic task, language used during exposure or memorization to stimuli can lead to nonlinguistic responses in alignment with language use.

Nonlinguistic Subject Variables

Even more heterogonous to a subject sample than the linguistic variables are the cultural, educational, and other experiential variables. Subject questionnaires are the usual ways to try to control these variables in post hoc analysis, but this control is limited by the foresight to collect adequate information.

One of the more obvious variables to control or record is the amount of schooling and literacy. Unfortunately, while schooling is easily represented on an ordinal scale (first to postsecondary grades), there is little guarantee that this represents the same education especially across, but even within, two population samples. For example, literacy is also not as simple a variable as it might appear. Subjects may be literate in different languages (and scripts) and may have very different literacy practices. Coding subjects who only read the Bible in their nonnative language and other subjects who read a variety of materials in their native language as both simply "literate" clearly glosses over potentially significant differences in experience.

Expertise may also vary considerably across samples. One of the most thorny obstacles in cross-cultural psychology is comparing testing results across two populations, one of which habitually engages with experiment-like settings and the other of which does not. This may have effects beyond simple difficulty in performance, but may affect the way in which subjects understand instructions, second-guess the intentions of the experimenter, and so on.[4]

Sex or gender, age, and the more physiologically based experiences are also difficult to compare. Being a woman in different societies means very different daily experiences beyond the variables of amount of schooling and the like. To what extent are subjects in their thirties the same across two populations. In one society but not another, a 35-year-old might typically be a grandparent in declining health with uncorrected vision or hearing loss.

Testing Environment

Lastly, variation in the testing environment is often difficult to control. The more broadly cross-cultural the samplings, the greater the dependence on local conditions. One might think of the ideal as an identical laboratory setup for each population sampled. However, since different subjects might react differently within such an environment, this is not necessarily a panacea (in addition to the obvious practical difficulty in implementation).

The best approach is to carefully examine the environmental features needed for the task at hand. If an experiment is about color categorization, lighting obviously needs to be controlled; if an experiment is about spatial arrays, adjacent landmarks and handedness need to be controlled; and so on. For example, in the basic experiment reported in Pederson et al. (1998), the use of table tops was not considered essential for tasks testing "table-top space," but the use of two delimited testing surfaces and the geometrical relationship and distances between these surfaces was critical to the design. This allowed the individual experimenters to set up tables or mats on the ground/floor as was more appropriate for the broader material culture.[5]

3.8. Establishing Causal Directionality

Once a correlation between a language pattern and a behavioral response has been experimentally established, the problem of establishing causal directionality remains. While this is a problem for any correlational design, it is particularly vexing for studies of linguistic relativity. Quite simply, it is difficult to rule out the possibility that subjects habitually speak the way they do as a consequence of their culture (and environment) as opposed to the possibility that the culture thinks the way it does because of their language. For discussions of the role of culture vis-à-vis language in linguistic relativity studies, see Bickel (2000), Enfield (2000), and the fairly standard reference of Hanks (1990).

In specific response to work on spatial cognition, Li and Gleitman (2002) argue that behavioral response patterns are not causally attributable to community language preferences, but rather that language use reflects cultural practice and concerns, for example, the many words for snow used by skiers—however, see also Levinson et al. (2002) for an extensive response. To the extent that the language features under investigation are roughly as changeable as the culture, this is certainly a likely possibility. On the other hand, when the language features are essentially fossilized in the grammatical system, they cannot be understood as the consequences of current cultural conditions. If anything, the pattern of grammaticized distinctions reflects the fossilized conceptualizations of one's ancestors.

4. Work within Cognitive Linguistics

Some of the earliest cognitive linguistic work (1970s) explicitly tying grammatical structure to cognition is found in studies by Talmy (see especially Talmy 1977, 1978). This work largely focuses on the universal (or at least broadly found) patterns of language and has been revised and expanded in Talmy (2000a, 2000b). Talmy treats language as one of many "cognitive systems" which has the "set of grammatically specified notions [constitute] the fundamental conceptual structuring system of language. . . . Thus, grammar broadly conceived, is the determinant of conceptual structure within one cognitive system, language" (2000a: 21–22). However, the relationship between this cognitive system (language) and others (i.e., nonlinguistic cognition) is relatively unspecified in his work. Structural commonalities between the various cognitive systems are suggested—most specifically between visual perception and language—but, importantly, Talmy avoids claims that there is any causal effect from linguistic categories to nonlinguistic categories.[6]

Langacker is bolder in the relationship between grammar and cognition: in Cognitive Grammar's "view of linguistic semantics. Meaning is equated with

conceptualization (in the broadest sense)" (Langacker 1987: 55). Langacker (1991) further argues that the cognitive models underlying clause structure have prototypes which are rooted in (variable) cultural understanding. To the extent that we find interesting cross-linguistic variation, we can see the work of Talmy and Langacker as sources for linguistic relativity hypotheses to test—as, for example, Slobin (1996, 2000) has begun with the motion event typology of Talmy (1985).

As mentioned above, Lakoff (1987: chapter 18) directly addresses linguistic relativity. Within this chapter on linguistic relativity, there is a discussion of different ways in which two cross-linguistic systems might be "commensurate." They might be *translatable*, *understandable* (though this is vaguely defined), *commensurate* in usage, share the same *framing*, and/or use the same *organization* of the various underlying concepts. In addition to a summary of the now classic Kay and Kempton (1984), there is an elaborate extension to linguistic relativity of semantics work in Mixtec and English by Brugman (1981) and Brugman and Macaulay (1986).

Metaphor is an obvious area of interest to many cognitive linguists (see Grady, this volume, chapter 8, and references therein). The nature of metaphor is to consider conceptualizations in terms of other linguistically expressed domains. To the extent that source domains can vary cross-linguistically or cross-culturally (or different features of these source domains are mapped), this is an area ripe for linguistic relativity studies. To date, however, linguistic relativity studies—that is to say, work with behavioral data—have largely limited themselves to the study of elemental and literal language. One exception to this is linguistic relativity research on time, which almost necessarily is metaphorically expressed (see section 5.6 below).

5. Research by Topic Area

This section gives a brief overview of modern linguistic relativity work organized by topic area. While some comments are given, it is impossible in this space to summarize the findings of the entire body of work. Further, the empirical details of each study are essential to critical evaluation of the findings, so the original sources must be consulted.

5.1. Color

Perhaps the greatest debate in linguistic relativity has been in the domain of color. Historically, linguists and anthropologists had been struck by the seemingly boundless diversity in color nomenclature. Given the obvious biological underpinnings of color perception, this made "color" a domain of choice to seek language-specific effects overriding biological prerequisites.

Lenneberg and Roberts (1956) is one of the earliest attempts to empirically test linguistic relativity, and as such this study spends considerable space defining the intellectual concerns before it reports on a relatively small study involving Zuni versus English color categorization. Brown and Lenneberg (1958) report on various work and develop the notion of *codability*: that is, the use of language as a way to more efficient coding of categories for the purposes not only of communication, but also of augmenting personal memory.

Berlin and Kay (1969) and the updated methodology in Kay and McDaniel (1978) have laid the groundwork of considerable research in color terminology. Central to the method is the use of Munsell color chips as a reference standard which can be carried to various field sites. Universal patterns were found to establish a typology of different color systems which appeared to be built out of a small set of universal principles. Research continues to be robust in this area and the interested reader may wish to consult the conference proceedings published as Hardin and Maffi (1997) for more current perspectives.

Eleanor Rosch (under her previous name: Heider 1971, 1972) found that focal colors (or Hering primaries from Hering's theory of light and color, see Hering 1964) were better remembered even by young children and were also more perceptually salient for them. Further, Heider and Olivier (1972) and Rosch (1973) found that, even for members of a community (the Dani of Papua New Guinea) who had little color terminology at all, certain color examples were better remembered. She argues that these "natural" categories are generally favored in human learning and cognition. This work is often taken as support for universals of color perception, though since the Dani had no linguistic categories to sway them away from biologically primary colors, this cannot be taken as evidence against a potential linguistic influence on color perception.

The effects of language on color categorization could be seen in Kay and Kempton (1984), but any effects of language-specific color terms only surfaced under specific conditions, and the effects were not as robust as earlier researchers had hoped. Various proposals have been made to revise the Berlin and Kay approach in ways which accommodate linguistic relativity effects within a basically universally constrained system. Most notable of these is Vantage Theory, which seeks to explain multiple points of view—even within the putative universals of color perception—and how points of view may be linguistically mediated; see especially MacLaury (1991, 1995, 2000).

Work by Davies and colleagues has also expanded upon the work of Kay and Kempton (1984) by examining a variety of linguistic systems for denoting colors. They then test participants from these speech communities using various categorization tasks. For Turkish, see Oezgen and Davies (1998); for Setswana, English, and Russian, see Davies (1998), Davies and Corbett (1997), and Davies et al. (1998); see also Corbett and Davies (1997) for a discussion of method in language sampling for color terminology.

Especially within anthropology, there has been concern about the fundamental adequacy of the empirical method followed by Berlin and Kay (and later

modifications). Jameson and D'Andrade (1997) address the adequacy of the theory of color perception inherent in the use of the Munsell color system. Lucy (1997b) criticizes most work on color terminology as insufficiently descriptive of the actual linguistic properties of the color terms themselves: without an adequate investigation into these properties, it is unclear what the effects may be of forcing reference with these terms into the Munsell system. The worry is that the Munsell system will not only standardize the coding of the responses, but actually create standardized and unnatural responses rather than allowing the terms to refer to their actual reference.

For a survey of recent work exploring color naming and its relationship to nonlinguistic cognition, see Kay and Regier (2006).

5.2. Shape Classification

In determining whether or not the Navajo shape classification system influenced sorting behavior, Carroll and Casagrande (1958) attempted to balance cultural factors across samples by using English-speaking and Navajo-speaking ethnic Navajo children. As a control group, English-speaking, middle-class American children were used. The results from triad classification (by either shape/function or color) were largely consistent with the Navajo verb classification, in that the Navajo-speaking Navajo children demonstrated a greater preference for shape sorting than English-speaking Navajo children. Note, however, that English-speaking middle-class children also patterned like Navajo-speaking children, suggesting to Carroll and Casagrande that cultural factors beyond language play an important role in such classification.

Lucy and Gaskins (2001) also use triad-type methods to compare Yucatecan children and adults with English-speaking Americans. Again, a broad consistency with each language's classification system is found, but interestingly, this only becomes prominent after age nine (see section 5.6)

5.3. Conditional Reasoning

With basic reasoning processes, variation is more likely to be viewed as directly advantageous or disadvantageous, that is, essentially correct or incorrect. Whether the hypothesized cause is linguistic or otherwise, in modern academia, the burden of proof appropriately falls most heavily on the researcher hoping to demonstrate any potential absence (or "deficiency") within a particular community.

The work of Alfred Bloom and his many detractors falls fully into this predicament. Bloom (1981) proposed that Chinese (unlike English) lacks a specific counterfactual construction and that this has led to reduced ability to engage in counterfactual reasoning. The debate was carried across several volumes of *Cognition*: Au (1983, 1984), Bloom (1984), Liu (1985), Takano (1989); making use of

different samples, these studies did not generally replicate Bloom's findings.[7] Unfortunately, there has been a tendency to interpret the various results (or lack thereof) as disconfirming linguistic relativity more generally rather than demonstrating a failure of a particular experimental design. Takano used Japanese speakers, who like Chinese speakers, lack a *dedicated* counterfactual construction, but found that their reasoning patterned like English speakers. More recently, Lardiere (1992) investigated Arabic speakers. Arabic patterns like English in that there is an explicit counterfactual construction, yet the Arabic participants performed like Bloom's original Chinese subjects on counterfactual reasoning. From these studies, both Takano and Lardiere conclude that the principal effect on counterfactual reasoning is traceable not to linguistic habit, but to cultural practices of reasoning, testing conventions, and the like.

Another conclusion one might draw from these studies is that we cannot automatically assume that either linguistic or nonlinguistic habit will be discernable from the presence or absence of specialized linguistic constructions. Obviously, those Chinese and Japanese speakers trained in formal counterfactual reasoning must have found some means of expression. Conversely, the Arabic speakers need not have used their counterfactual construction in ways analogous to the ways of formally educated English speakers.

5.4. Number

Cardinal Numbers

One clear way in which languages vary is in their cardinal number systems. In addition to the obvious lack of larger numbers in many languages (at least as native vocabulary), languages also vary in their organization of these numbers. Various languages partially use a base twenty counting system and other languages appear to have relics of base five systems. But even within primarily base ten systems, there is variation of consistency and expression.

Miura (1987) argues that the generally superior mathematical abilities of school children in or from some cultures (especially East Asian) result at least in part from the transparency and exception-free nature of the base ten numerals used for counting, which children generally control prior to beginning formal education—see also the follow-up cross-linguistic studies: Miura and Okamoto (1989), Miura et al. (1988), Miura et al. (1993), Miura et al. (1994), Miura et al. (1999).

Saxton and Towse (1998) provide a more cautious conclusion, suggesting that the influence of native language on the task of learning place values is less than argued for by Miura and colleagues. Many other differences in performance were found across groups which were better accounted for as resulting from general cultural attitudes toward education and so on, than as the result of the linguistic number system.

Grammatical Number

On a grammatical level, languages vary in terms of their grammatical encoding of the number of entities in an event or scene. While this topic has not been widely taken up, the work of Lucy (1992a) is noteworthy for its extensive consideration of attention to number in Mayan and English speakers. An extensive typological discussion of grammatical number, though without focus on issues of linguistic relativity, is provided by Corbett (2000). Lastly, Hill and Hill (1998) discuss the effects of culture on language (rather than linguistic relativity) for number marking (plurals), and in particular the "anti-Whorfian effect" they find in Uto-Aztecan.

5.5. Space

Reference Frames

Currently, the primary area of linguistic relativity research in spatial domains is with reference frames (however, there is also the important developmental work on topological relations by Choi and Bowerman 1991, see below).

Reference frames are the psychological or linguistic representation of relationships between entities in space. They require fixed points of reference, such as the speaker, a landmark, or an established direction. Within linguistics, the typology of reference frames is complicated, but most accounts include something like an intrinsic reference frame (whereby an object is located only with respect to an immediate point, e.g., *The ball is next to the chair*) and various flavors of reference frames which make use of additional orientation (e.g., *The ball is to my right of the chair* or *The ball is to the north of the chair*). Languages vary in terms of their habitually selected reference frames, and following the linguistic relativity hypothesis, speakers should also vary in their encoding spatial memories, making locational calculations, and so forth. For extensive work measuring event-related potential data (recordings at the scalp of electrical charges from brain activity during specific tasks), see the work of Taylor and colleagues: Taylor et al. (1999) and Taylor et al. (2001). These works compare the viewer/speaker-relative (or *egocentric*) reference frame with the intrinsic.

Of note for being broadly comparative across diverse linguistic and cultural communities is the work reported in Pederson et al. (1998), which found correlations between habitual linguistic selection of reference frames and cognitive performance on spatial memory (and other) tasks. There were many studies within this same general project. Perhaps the most important to consult for the theoretical underpinnings for the project are Brown and Levinson (1993) and Levinson (1996). As pointed out by Li and Gleitman (2002), the populations reported as using an absolute/geocardinal (*north of . . .*) reference frame were largely rural populations, and the populations using a speaker-relative/egocentric reference frame are largely urban, so there is a potential confound in the population samples between language

and culture/environment. For a rebuttal to these concerns and Li and Gleitman's similar experiments, see Levinson et al. (2002); see also Pederson (1998) for a discussion of this urban/rural cultural split.

Motion Events

Talmy (1985, 2000b) identifies a typological contrast in the ways that languages encode basic motion events. To simplify, some languages such as the Romance languages commonly encode the fact of motion and the basic path with the main verb (e.g., *to enter*, *to ascend*, etc.). In contrast, Germanic and many other languages most commonly encode the fact of motion along with the manner of motion in the verb (e.g., *to wiggle*), and the path is expressed elsewhere.

Slobin (1991, 1996) considers the cognitive consequences of these linguistic patterns for English and Spanish speakers. Slobin (2000) extends this approach to French, Hebrew, Russian, and Turkish. Gennari et al. (2002) and Malt, Sloman, and Gennari (2003) examine these contrasts experimentally and argue for some effects of one's native language pattern on certain nonlinguistic tasks.

5.6. Time

While spatial relationships have been extensively studied for linguistic relativity effects, the effects of different temporal encoding have received much less attention. In part, this may be attributed to the relative difficulty of developing research instruments. An obvious difference cross-linguistically is whether or not a language grammatically encodes tense. Bohnemeyer (1998) discusses the lack of tense-denoting constructions in Yucatec Mayan and contrasts this with German speakers observing the same video stimuli; nonetheless, both samples appeared to have encoded similar event orderings in memory. Languages also have some variation in preferred metaphors for talking about time. Boroditsky (2000, 2001) argues that Mandarin Chinese speakers have a different metaphor for time (vertical) and this appears to influence their nonlinguistic encoding as well.

5.7. Developmental Studies

Ultimately, any linguistic relativity effects must be explained in terms of the acquisition of linguistic categories and the effects on cognitive development.

Choi and Bowerman (1991) and Bowerman and Choi (2001) contrast early lexical acquisition of Korean and English spatial terms, principally those expressing contact, closure, and similar concepts. Korean-speaking adults use spatial terms to categorize subtypes of these different relationships in very different ways from English-speaking adults. Perhaps surprisingly, Choi and Bowerman report that Korean-speaking children as young as two demonstrate linguistic patterning more like the Korean-speaking adults than like the English-speaking children (and vice versa). This suggests that even in fairly early lexical acquisition, children show remarkable sensitivity

to the specific language input rather than relying on purportedly universal cognitive categorizations and fitting the language categories onto these.

Lowenstein and Gentner (1998), Gentner and Loewenstein (2002), and Gentner and Boroditsky (2001) argue that metaphor and analogical reasoning are key parts of concept development and early word meaning. To the extent that these are cross-linguistically variable, it can be argued that linguistic relativity effects may be present especially for abstract reasoning which most depends on relational terminology and analogy.

As mentioned in the section on shape classification, Lucy and Gaskins (2001) look at the age of development of language-particular patterns in shape versus material sorting tasks. Assuming one can extrapolate from their data, the critical age at which language helps to direct nonlinguistic behavior (for these sorts of tasks) is around ages 7–9. This suggests that the acquisition of language categories need not immediately manifest cognitive effects in nonlinguistic domains, but rather that there may be a period in which the linguistic categories are initially more solely linguistic and then eventually the analogy from language to other types of categorization is drawn. It may also reflect a greater dependence on linguistically mediated internal thought, à la Vygotsky.

Susan Goldin-Meadow and colleagues have examined the interplay of gesture, home sign, and conventional language use and their relationships to underlying (and developing) cognitive representations. A good recent summary may be found in Goldin-Meadow (2002) and the references within. Zheng and Goldin-Meadow (2002) examine the similarities across cultures in home sign despite notable differences in the adult spoken languages. These commonalities suggest what the underlying conceptual categories may be in children prior to acquiring the "filter" provided by the model of a specific language.

Working with English-speaking children and language acquisition delayed deaf children, de Villiers and de Villiers (2000) argue that language has a vital role in the development of understandings of false beliefs—at least insofar as demonstrated in unseen displacement. (For example, the puppet doesn't see that I replaced the crayons in the crayon box with a key; what does the puppet think is in the crayon box?) Language is eminently suited for the representation of counterfactual and alternative beliefs, so it is unclear whether it is the specifics of language acquisition or just general exposure to alternatives that happen to come through the medium of language which might be driving this development. For a summary of the work by Gopnik and colleagues on the potential interactions of language and cognitive development, especially around ages 1–2, see Gopnik (2001).

5.8. Sign Language versus Spoken Language

Lastly, what of the medium of the language itself? Might the mechanical constraints of spoken language versus sign language have their own influences? Working with native ASL signers and English speakers on mental rotation tasks,

Emmorey, Klima, and Hickok (1998) show evidence that the vast experience of signers in understanding their interlocutors' spatial perspective during signing has given them some advantage in nonlinguistic rotation tasks compared with nonsigners.

6. Future Directions

As can be seen from the above discussion, the issue of linguistic relativity is as open a question as it is broad. However, as empirically driven models of human cognition become increasingly detailed, work within linguistic relativity (and Cognitive Linguistics generally) becomes increasingly specific in its description of cognitive mechanisms.

The question "Does language influence thought?" is being replaced by a battery of questions about whether a given feature of a specific language influences particular cognitive operations, what the exact cognitive mechanisms are which give rise to this influence, and how we can most precisely characterize the nature of this influence? Rather than this being a step away from the "big picture" of human cognition, this general trend toward increasingly precise definitions and, ideally, more falsifiable hypotheses leads us to a simply more reliable understanding of cognition and the role of language within it.

As we discover more of the specific interactions between language and the rest of the cognitive systems, there is a need to understand the time course of this development. Except for Lucy and Gaskins (2001) and some of the home sign studies, there has been virtually no attempt to determine the time course of any linguistic relativity effects. If language influences a particular cognitive operation or conceptualization, does it do so upon acquisition of the language model, shortly subsequent to this acquisition, or is there a gradual "internalization" (in Vygotskian terms) of the linguistic structure as something more than a learned code?

One must also wonder whether certain linguistic construals more readily have influences beyond language than others. For example, is spatial categorization more likely to be influenced by language than color categorization is, or vice versa? If some domains are more linguistically sensitive, what do these domains have in common?

These are all broad questions and are unlikely to be resolved in the immediate future. However, as research in linguistic relativity becomes increasingly mainstream within psychology and linguistics, it seems certain that we will understand ever more of the complexities between language and thought.

NOTES

1. Many more recent writings by Alford on Whorf, linguistic relativity, and related topics can be found on Alford's Web site: http://www.enformy.com/alford.htm.

2. This idea was apparently insufficiently discredited as it has more recently resurfaced in the popular press with Shlain (1998)—where it is now associated with the demise of polytheism and the claimed consequent surge of misogyny in European history.

3. Anecdotally, I can report that subjects in spatial reference frame experiments would use their linguistically dominant frame of reference in nonlinguistic tasks but would switch when they heard an alternate frame of reference used immediately before the task. (Specifically, when an assistant erroneously used nonneutral language in an example.) In subsequent tasks, with no reference frame language repeated, the subjects could switch over to what might well have been a more default reference frame for such tasks. Of course, these subject results are not coded with other subjects, and this dictated extreme care in controlling the immediately preceding linguistic environment during experimental sessions.

4. College students (especially those participating for credit in an introductory psychology class!) are infamous for trying to second guess the "hidden" purpose of an experiment. Surely, such subjects are less directly comparable with the perhaps experimentally less savvy subjects drawn from other populations.

5. Li and Gleitman (2002) changed "small procedural details" (see their footnote 5) in this experiment—notably they eliminated the distance between the tables—and report different results. Although they do not attribute the different results to these changes, but rather to other uncontrolled variables in the original study, the control of the experimental setup clearly can be critical for evaluating the results.

6. The linguistic parallels with basic operations in visual perception imply a bias favoring the building of linguistic categories from more fundamental cognitive categories rather than any particular influence from language to cognition.

7. Cara and Politzer (1993) also found no correspondence of language to reasoning with Chinese and English speakers on counterfactual reasoning tasks, though the design seems uninfluenced by the debate in *Cognition*.

REFERENCES

Alford, Dan K. H. 1978. The demise of the Whorf hypothesis. *Berkeley Linguistics Society* 14: 485–99.

Au, Terry Kit-Fong. 1983. Chinese and English counterfactuals: The Sapir-Whorf hypothesis revisited. *Cognition* 15: 155–87.

Au, Terry Kit-Fong. 1984. Counterfactuals: In reply to Alfred Bloom. *Cognition* 17: 289–302.

Berlin, Brent. 1978. Ethnobiological classification. In Eleanor Rosch and Barbara B. Lloyd, eds., *Cognition and categorization* 9–26. Hillsdale, NJ: Lawrence Erlbaum.

Berlin, Brent, Dennis E. Breedlove, and Peter H. Raven. 1973. General principles of classification and nomenclature in folk biology. *American Anthropologist* 75: 214–42.

Berlin, Brent, and Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bickel, Balthasar. 2000. Grammar and social practice: On the role of 'culture' in linguistic relativity. In Susanne Niemeier and René Dirven, eds., *Evidence for linguistic relativity* 161–91. Amsterdam: John Benjamins.

Bloom, Alfred H. 1981. *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West*. Hillsdale, NJ: Lawrence Erlbaum.

Bloom, Alfred H. 1984. Caution—the words you use may affect what you say: A response to Au. *Cognition* 17: 275–87.

Blount, Ben G. 1993. Cultural bases of folk classification systems. In Jeanette Altarriba, ed., *Cognition and culture: A cross-cultural approach to psychology* 3–22. Amsterdam: Elsevier.

Bohnemeyer, Jürgen. 1998. Time relations in discourse: Evidence from a comparative approach to Yukatek Maya. PhD dissertation, Katholieke Universiteit Brabant, Netherlands.

Boroditsky, Lera. 2000. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* 75: 1–28.

Boroditsky, Lera. 2001. Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology* 43: 1–22.

Bowerman, Melissa, and Soonja Choi. 2001. Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In Melissa Bowerman and Stephen C. Levinson, eds., *Language acquisition and conceptual development* 475–511. Cambridge: Cambridge University Press.

Brown, Penelope, and Stephen C. Levinson. 1993. "Uphill" and "downhill" in Tzeltal. *Journal of Linguistic Anthropology* 3: 46–74.

Brown, Roger Langham. 1967. *Wilhelm von Humboldt's conception of linguistic relativity: Janua linguarum*. The Hague: Mouton.

Brown, Roger W., and Eric. H. Lenneberg. 1958. Studies in linguistic relativity. In Eleanor E. Maccoby, Theodore M. Newcomb, and Eugene L. Hartley, eds., *Readings in social psychology* 9–18. New York: Holt, Rinehart and Winston.

Brugman, Claudia. 1981. Story of *Over*. MA thesis, University of California at Berkeley. (Published as *The story of Over: Polysemy, semantics, and the structure of the lexicon*. New York: Garland, 1988)

Brugman, Claudia, and Monica Macaulay. 1986. Interacting semantic systems: Mixtec expressions of location. *Berkeley Linguistics Society* 12: 315–27.

Cara, Francesco, and Guy Politzer. 1993. A comparison of conditional reasoning in English and Chinese. In Jeanette Altarriba, ed., *Cognition and culture: A cross-cultural approach to psychology* 283–97. Amsterdam: Elsevier.

Carroll, John B., and Joseph B. Casagrande. 1958. The function of language classifications in behavior. In Eleanor. E. Maccoby, Theodore M. Newcomb, and Eugene L. Hartley, eds., *Readings in social psychology* 18–31. New York: Holt, Rinehart and Winston.

Carter, Anthony T. 1973. A comparative analysis of systems of kinship and marriage in South Asia. *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland*: 29–54.

Chafe, Wallace. 2000. Loci of diversity and convergence in thought and language. In Martin Pütz and Marjolijn Verspoor, eds., *Explorations in linguistic relativity* 101–23. Amsterdam: John Benjamins.

Choi, Soonja, and Melissa Bowerman. 1991. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* 41: 83–121.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Corbett, Greville G., and Ian R. L. Davies. 1997. Establishing basic color terms: Measures and techniques. In C. L. Hardin and Luisa Maffi, eds., *Color categories in thought and language* 197–223. Cambridge: Cambridge University Press.

Danziger, Eve, and Eric Pederson. 1998. Through the looking glass: Literacy, writing systems and mirror image discrimination. *Written Language and Literacy* 1: 153–64.

Davies, Ian R. L. 1998. A study of colour grouping in three languages: A test of the linguistic relativity hypothesis. *British Journal of Psychology* 89: 433–52.

Davies, Ian R. L., and Greville G. Corbett. 1997. A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology* 88: 493–517.

Davies, Ian R. L., Paul T. Sowden, David T. Jerrett, Tiny Jerrett, and Greville G. Corbett. 1998. A cross-cultural study of English and Setswana speakers on a colour triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology* 89: 1–15.

de Villiers, Jill G., and Peter A. de Villiers. 2000. Linguistic determinism and the understanding of false beliefs. In Peter Mitchell and Kevin John Riggs, eds., *Children's reasoning and the mind* 191–228. East Sussex: Psychology Press.

Duranti, Alessandro. 1997. *Linguistic anthropology*. Cambridge: Cambridge University Press.

Emmorey, Karen, Edward Klima, and Gregory Hickok. 1998. Mental rotation within linguistic and non-linguistic domains in users of American sign language. *Cognition* 68: 221–46.

Enfield, Nick J. 2000. On linguocentrism. In Martin Pütz and Marjolijn Verspoor, eds., *Explorations in linguistic relativity* 125–57. Amsterdam: John Benjamins.

Foley, William A. 1997. *Anthropological linguistics: An introduction*. Oxford: Basil Blackwell.

Geeraerts, Dirk, Stefan Grondelaers, and Peter Bakema. 1994. *The structure of lexical variation: Meaning, naming, and context*. Berlin: Mouton de Gruyter.

Gennari, Silvia P., Steven A. Sloman, Barbara C. Malt, and W. Tecumseh Fitch. 2002. Motion events in language and cognition. *Cognition* 83: 49–79.

Gentner, Dedre, and Lera Boroditsky. 2001. Individuation, relativity, and early word learning. In Melissa Bowerman and Stephen C. Levinson, eds., *Language acquisition and conceptual development* 215–56. Cambridge: Cambridge University Press.

Gentner, Dedre, and Jeffrey Loewenstein. 2002. Relational language and relational thought: Language, literacy, and cognitive development. In Eric Amsel and James P. Byrnes, eds., *The development and consequences of symbolic communication* 87–120. Mahwah, NJ: Lawrence Erlbaum.

Goldin-Meadow, Susan. 2002. From thought to hand: Structured and unstructured communication outside of conventional language. In Eric Amsel and James P. Byrnes, eds., *Language, literacy, and cognitive development: The development and consequences of symbolic communication* 121–50. Mahwah, NJ: Lawrence Erlbaum.

Goody, Jack, and Ian Watt. 1962. The consequences of literacy. *Comparative Studies in Sociology and History* 5: 304–45.

Gopnik, Alison. 2001. Theories, language and culture: Whorf without wincing. In Melissa Bowerman and Stephen C. Levinson, eds., *Language acquisition and conceptual development* 45–69. Cambridge: Cambridge University Press.

Graff, Harvey J. 1987. *The legacies of literacy*. Bloomington: Indiana University Press.

Hanks, William F. 1990. *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.

Hardin, C. L., and Luisa Maffi. 1997. *Color categories in thought and language*. Cambridge: Cambridge University Press.

Heider, Eleanor R. 1971. "Focal" color areas and the development of color names. *Developmental Psychology* 4: 447–55.
Heider, Eleanor R. 1972. Universals in color naming and memory. *Journal of Experimental Psychology* 93: 10–20.
Heider, Eleanor R., and Donald C. Olivier. 1972. The structure of the color space in naming and memory in two languages. *Cognitive Psychology* 3: 337–54.
Hering, Ewald. 1964. *Outlines of a theory of the light sense*. Cambridge, MA: Harvard University Press.
Hill, Jane H., and Kenneth C. Hill. 1998. Culture influencing language: Plurals of Hopi kin terms in comparative Uto-Aztecan perspective. *Journal of Linguistic Anthropology* 7: 166–80.
Hill, Jane H., and Bruce Mannheim. 1992. Language and world view. *Annual Review of Anthropology* 21: 381–406.
House, Juliane. 2000. Linguistic relativity and translation. In Martin Pütz and Marjolijn Verspoor, eds., *Explorations in linguistic relativity* 69–88. Amsterdam: John Benjamins.
Hunn, Eugene S. 1977. *Tzeltal folk zoology: The classification of discontinuities in nature*. New York: Academic Press.
Hunn, Eugene S. 1982. The utilitarian factor in folk biological classification. *American Anthropologist* 84: 830–47.
Hunt, Earl, and Franca Agnoli. 1991. The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review* 98: 377–89.
Jameson, Kimberly, and Roy G. D'Andrade. 1997. It's not really red, green, yellow, blue: An inquiry into perceptual color space. In C. L. Hardin and Luisa Maffi, eds., *Color categories in thought and language* 295–319. Cambridge: Cambridge University Press.
Kay, Paul. 1996. Intra-speaker relativity. In John J. Gumperz and Stephen C. Levinson, eds., *Rethinking linguistic relativity* 97–114. Cambridge: Cambridge University Press.
Kay, Paul, and Willett Kempton. 1984. What is the Sapir-Whorf hypothesis? *American Anthropologist* 86: 65–79.
Kay, Paul, and Chad K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. *Language* 54: 610–46.
Kay, Paul, and Terry Regier. 2006. Language, thought and color: Recent developments. *Trends in Cognitive Sciences* 10: 51–54.
Koerner, Konrad E. F. 2000. Towards a 'full pedigree' of the 'Sapir-Whorf Hypothesis': From Locke to Lucy. In Martin Pütz and Marjolijn Verspoor, eds., *Explorations in linguistic relativity* 1–23. Amsterdam: John Benjamins.
Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Langacker, Ronald W. 1987. Nouns and verbs. *Language* 63: 53–94.
Langacker, Ronald W. 1991. *Foundations of cognitive grammar*. Vol. 2, *Descriptive application*. Stanford, CA: Stanford University Press.
Lardiere, Donna. 1992. On the linguistic shaping of thought: Another response to Alfred Bloom. *Language in Society* 21: 231–51.
Lee, Penny. 1996. *The Whorf theory complex: A critical reconstruction*. Amsterdam: John Benjamins.
Lee, Penny. 2000. When is 'linguistic relativity' Whorf's linguistic relativity? In Martin Pütz and Marjolijn Verspoor, eds., *Explorations in linguistic relativity* 45–68. Amsterdam: John Benjamins.

Lenneberg, Eric H., and John M. Roberts. 1956. *The language of experience: A study in methodology*. International Journal of American Linguistics Memoir, no. 13. Baltimore, MD: Waverly Press.

Levinson, Stephen C. 1996. Frames of references and Molyneux's question: Crosslinguistic evidence. In Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merrill Garrett, eds., *Language and space* 109–69. Cambridge, MA: MIT Press.

Levinson, Stephen C., Sotaro Kita, Daniel B. M. Hauna, and Björn H. Rasch. 2002. Returning the tables: Language affects spatial reasoning. *Cognition* 84: 155–88.

Li, Peggy, and Lila Gleitman. 2002. Turning the tables: Language and spatial reasoning. *Cognition* 83: 265–94.

Liu, Lisa Gabern. 1985. Reasoning counterfactually in Chinese: Are there any obstacles? *Cognition* 21: 239–70.

Loftus, Elizabeth F. 1975. Leading questions and the eyewitness report. *Cognitive Psychology* 7: 560–72.

Losonsky, Michael, ed. 1999. *Wilhelm von Humboldt: On language: On the diversity of human language construction and its influence on the mental development of the human species*. Cambridge: Cambridge University Press.

Lowenstein, Jeffrey, and Dedre Gentner. 1998. Relational language facilitates analogy in children. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* 615–20. Mahwah, NJ.: Lawrence Erlbaum.

Lucy, John A. 1992a. *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, John A. 1992b. *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, John A. 1997a. Linguistic relativity. *Annual Review of Anthropology* 26: 291–312.

Lucy, John A. 1997b. The linguistics of "color." In C. L. Hardin and Luisa Maffi, eds., *Color categories in thought and language* 320–46. Cambridge: Cambridge University Press.

Lucy, John A., and Suzanne Gaskins. 2001. Grammatical categories and the development of classification preferences: A comparative approach. In Melissa Bowerman and Stephen C. Levinson, eds., *Language acquisition and conceptual development* 257–83. Cambridge: Cambridge University Press.

Lucy, John A., and Richard A. Shwedder. 1988. The effect of incidental conversation on memory for focal colors. *American Anthropologist* 90: 923–31.

Luria, Aleksandr Romanovich. 1976. *Cognitive development, its cultural and social foundations*. Cambridge, MA: Harvard University Press.

MacLaury, Robert E. 1991. Exotic color categories: Linguistic relativity to what extent? *Journal of Linguistic Anthropology* 1: 26–51.

MacLaury, Robert E. 1995. Vantage theory. In John R. Taylor and Robert E. MacLaury, eds., *Language and the cognitive construal of the world* 231–76. Berlin: Mouton de Gruyter.

MacLaury, Robert E. 2000. Linguistic relativity and the plasticity of categorization: Universalism in a new key. In Martin Pütz and Marjolijn H. Verspoor, eds., *Explorations in linguistic relativity* 251–93. Amsterdam: John Benjamins.

Malt, Barbara C., Steven A. Sloman, and Silvia P. Gennari. 2003. Speaking versus thinking about objects and actions. In Dedre Gentner and Susan Goldin-Meadow, eds., *Language in mind: Advances in the study of language and thought* 81–111. Cambridge, MA: MIT Press.

Manchester, Martin L. 1985. *The philosophical foundations of Humboldt's linguistic doctrines*. Amsterdam: John Benjamins.

Miura, Irene T. 1987. Mathematics achievement as a function of language. *Journal of Educational Psychology* 79: 79–82.

Miura, Irene T., Chungsoon C. Kim, Chih-mei Chang, and Yukari Okamoto. 1988. Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development* 59: 1445–50.

Miura, Irene T., and Yukari Okamoto. 1989. Comparisons of U.S. and Japanese first graders' cognitive representation of number and understanding of place value. *Journal of Educational Psychology* 81: 109–14.

Miura, Irene T., Yukari Okamoto, Chungsoon C. Kim, Chih-Mei Chang, Marcia Steere, and Michel Fayol. 1994. Comparisons of children's cognitive representation of number: China, France, Japan, Korea, Sweden, and the United States. *International Journal of Behavioral Development* 17: 401–11.

Miura, Irene T., Yukari Okamoto, Chungsoon C. Kim, Marcia Steere, and Michel Fayol. 1993. First graders' cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology* 85: 24–30.

Miura, Irene T., Yukari Okamoto, Vesna Vlahovic-Stetic, Chungsoon C. Kim, and Jong Hye Han. 1999. Language supports for children's understanding of numerical fractions: Cross-national comparisons. *Journal of Experimental Child Psychology* 74: 356–65.

Morais, José, Luz Cary, Jésus Alegria, and Paul Bertelson. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7: 323–31.

Oezgen, Emre, and Ian R. L. Davies. 1998. Turkish color terms: Tests of Berlin and Kay's theory of color universals and linguistic relativity. *Linguistics* 36: 919–56.

Olson, David R. 1991. Literacy as metalinguistic activity. In David R. Olson and Nancy Torrance, eds., *Literacy and orality* 251–70. Cambridge: Cambridge University Press.

Olson, David R. 2002. What writing does to the mind. In Eric Amsel and James P. Byrnes, eds., *Language, literacy, and cognitive development: The development and consequences of symbolic communication* 153–65. Mahwah, NJ: Lawrence Erlbaum.

Ong, Walter J. 1992. Writing is a technology that restructures thought. In Pam Downing, Susan D. Lima, and Michael Noonan, eds., *The linguistics of literacy* 293–319. Amsterdam: John Benjamins.

Pederson, Eric. 1993. Geographic and manipulable space in two Tamil linguistic systems. In Andrew U. Frank and Irene Campari, eds., *Spatial information theory* 294–311. Berlin: Springer Verlag.

Pederson, Eric. 1995. Language as context, language as means: Spatial cognition and habitual language use. *Cognitive Linguistics* 6: 33–62.

Pederson, Eric. 1998. Spatial language, reasoning, and variation across Tamil communities. In Petr Zima and Vladimír Tax, eds., *Language and location in space and time* 111–19. Munich: Lincom Europa.

Pederson, Eric. 2003. Mirror-image discrimination among nonliterate, monoliterate, and biliterate Tamil speakers. *Written Language and Literacy* 6: 71–91.

Pederson, Eric, Eve Danziger, David Wilkins, Stephen Levinson, Sotaro Kita, and Gunter Senft. 1998. Semantic typology and spatial conceptualization. *Language* 74: 557–89.

Read, Charles, Yun-Fei Zhang, Hong-Yin Nie, and Bao-Qing Ding. 1986. The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24: 31–44.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4: 328–50.
Saxton, Matthew, and John N. Towse. 1998. Linguistic relativity: The case of place value in multi-digit numbers. *Journal of Experimental Child Psychology* 69: 66–79.
Scinto, Leonard F. M. 1986. *Written language and psychological development*. New York: Academic Press.
Scribner, Sylvia, and Michael Cole. 1981. *The psychology of literacy*. Cambridge, MA: Harvard University Press.
Shlain, Leonard. 1998. *The alphabet versus the goddess: The conflict between word and image*. New York: Viking.
Silverstein, Michael. 1985. Language and the culture of gender: At the intersection of structure, usage, and ideology. In Elizabeth Mertz and Richard J. Parmentier, eds., *Semiotic mediation: Sociocultural and psychological perspectives* 219–59. Orlando, FL: Academic Press.
Silverstein, Michael. 1987. Cognitive implications of a referential hierarchy. In Maya Hickmann, ed., *Social and functional approaches to language* 125–64. Orlando, FL: Academic Press.
Slobin, Dan I. 1991. Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics* 1: 7–25.
Slobin, Dan I. 1996. Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibatani and Sandra A. Thompson, eds., *Grammatical constructions: Their form and meaning* 195–219. Oxford: Oxford University Press.
Slobin, Dan I. 2000. Verbalized events: A dynamic approach to linguistic relativity and determinism. In Susanne Niemeier and René Dirven, eds., *Evidence for linguistic relativity* 107–38. Amsterdam: John Benjamins.
Smith, Marion V. 1996. Linguistic relativity: On hypotheses and confusions. *Communication & Cognition* 29: 65–90.
Takano, Yohtaro. 1989. Methodological problems in cross-cultural studies of linguistic relativity. *Cognition* 31: 141–62.
Talmy, Leonard. 1977. Rubber-sheet cognition in language. *Chicago Linguistic Society* 13: 612–28.
Talmy, Leonard. 1978. The relation of grammar to cognition—a synopsis. In David Waltz ed., *Proceedings of TINLAP-2* 14–24. New York: Association for Computing Machinery.
Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical form. In Timothy Shopen, ed., *Language typology and syntactic description*, vol. 3, *Grammatical categories and the lexicon* 57–149. Cambridge: Cambridge University Press.
Talmy, Leonard. 2000a. *Toward a cognitive semantics*. Vol. 1, *Concept structuring systems*. Cambridge, MA: MIT Press.
Talmy, Leonard. 2000b. *Toward a cognitive semantics*. Vol. 2, *Typology and process in concept structuring*. Cambridge, MA: MIT Press.
Taylor, Holly A., Robert R. Faust, Tatiana Sitnikova, Susan J. Naylor, and Phillip J. Holcomb. 2001. Is the donut in front of the car? An electrophysiological study examining spatial reference frame processing. *Canadian Journal of Experimental Psychology* 55: 175–84.
Taylor, Holly A., Susan J. Naylor, Robert R. Faust, and Phillip J. Holcomb. 1999. "Could you hand me those keys on the right?" Disentangling spatial reference frames using different methodologies. *Spatial Cognition and Computation* 1: 381–97.
Whorf, Benjamin L. 1956. *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Whorf, Benjamin L., and George L. Trager. [1938] 1996. Report on linguistic research in the department of Anthropology of Yale University for the term Sept. 1937–June 1938. In Penny Lee, *The Whorf theory complex: A critical reconstruction* 251–80. Amsterdam: John Benjamins.

Zheng, Mingyu, and Susan Goldin-Meadow. 2002. Thought before language: How deaf and hearing children express motion events across cultures. *Cognition* 85: 145–74.