ELSEVIER

Cognition 84 (2002) 189–219

COGNITION

www.elsevier.com/locate/cognit

Shake, rattle, 'n' roll: the representation of motion in language and cognition

Anna Papafragou*, Christine Massey, Lila Gleitman

Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Suite 400A, Philadelphia, PA 19104, USA

Received 12 June 2001; received in revised form 12 December 2001; accepted 8 March 2002

Abstract

Languages vary strikingly in how they encode motion events. In some languages (e.g. English), manner of motion is typically encoded within the verb, while direction of motion information appears in modifiers. In other languages (e.g. Greek), the verb usually encodes the direction of motion, while the manner information is often omitted, or encoded in modifiers. We designed two studies to investigate whether these language-specific patterns affect speakers' reasoning about motion. We compared the performance of English and Greek children and adults (a) in nonlinguistic (memory and categorization) tasks involving motion events, and (b) in their linguistic descriptions of these same motion events. Even though the two linguistic groups differed significantly in terms of their linguistic preferences, their performance in the nonlinguistic tasks was identical. More surprisingly, the linguistic descriptions given by subjects *within language* also failed to correlate consistently with their memory and categorization performance in the relevant regards. For the domain studied, these results are consistent with the view that conceptual development and organization are largely independent of language-specific labeling practices. The discussion emphasizes that the necessarily sketchy nature of language use assures that it will be at best a crude index of thought.

Keywords: Motion; Spatial language; Spatial representation; Linguistic relativity

1. Introduction

Little within the topic of language so excites the popular imagination as the question of how the forms and content of a particular language might influence the thought of its users. Rightly, this question has come to be associated with Benjamin Whorf (1939, 1941) and

* Corresponding author. Fax: +1-215-573-9247.

E-mail addresses: anna4@linc.cis.upenn.edu (A. Papafragou), massey@linc.cis.upenn.edu (C. Massey), gleitman@cattell.psych.upenn.edu (L. Gleitman).

0010-0277/02/$ - see front matter
PII: S0010-0277(02)00046-X

Edward Sapir (1941), linguists who, in the first half of the 20th century, investigated the languages and cultural practices of several Native American tribes. As Whorf and Sapir understood, the finding that cultural differences are often mirrored in linguistic differences leaves all causal questions unresolved, but these commentators laid their bets on the languages themselves as the more significant engines of cultural and cognitive disparity. In Whorf's words:

> Language and culture are constantly influencing each other. But in this partnership the nature of the language is the factor that limits free plasticity and rigidifies channels of development in the more autocratic way… We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented as a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems in our minds. (Whorf, 1956, p. 213)

And relatedly, from Sapir:

> Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression. …"the real world" is to a large extent unconsciously built up on the language habits of the group. (Sapir, 1941, p. 75)

It is fair to say that "linguistic determinism" in the sense implied in these quotes virtually vanished from mainstream linguistic and psychological discussion for several decades (though in some weaker versions, often called "linguistic relativity", this tradition has continued to thrive within anthropological linguistics). The replacement approaches have been versions of a universalist perspective that emphasizes cross-linguistic similarities, assigning these to our shared human nature. Noam Chomsky is the leading modern proponent and explicator of this universalist approach to language studies. In the following passage, he presents this vision as perhaps the most serious reason for studying language at all:

> …Language is a mirror of mind in a deep and significant sense. It is a product of human intelligence… By studying the properties of natural languages, their structure, organization, and use, we may hope to learn something about human nature; something significant, if it is true that human cognitive capacity is the truly distinctive and most remarkable characteristic of the species. (Chomsky, 1975, p. 4)

These two perspectives on the psychology of language and thought could not be more different. Both camps express the passionate conviction that our language and thought are inextricably bound up together; but cause for one camp is effect for the other. The issue is by no means decided. While the universalist view propounded by Chomsky is perhaps still dominant in linguistic circles, there has been something of a Renaissance of Whorfian theorizing in recent psychological discussion. The issue, in fact, is worse than undecided: even the field of battle is in doubt, as investigators contend about where (in human perception and cognition) linguistic influences might be expected to surface. Worst of

all, the very question of language–thought relations is variously definable, with the consequence that present-day investigators struggle even to identify – no less adjudicate – several "stronger" and "weaker" interpretations of the relativist position.

1.1. Choosing the field of battle

The studies we will present in this paper represent an attempt to contribute to this debate by comparing the way that linguistic representations and processes do – and do not – impact and guide other cognitive functions; here, memory and categorization. Any such investigation is perforce topically limited. This raises the specter that we might require a virtual infinity of such studies lest by focusing on any single conceptual or cognitive domain one misses potential influences that, though invisible in the present case, may show up in the very next study that is done. Because one cannot investigate everything, then, there is considerable motivation for picking the ground carefully.

Two interlocking factors seem to have motivated this choice in the earliest psycholinguistic studies of the Sapir–Whorf hypothesis. First, of course, was that languages show variation on the chosen dimension. Second, there had to be some objective way to measure discrimination for the studied function. An obvious choice was *color* (usually *hue* but sometimes also *brightness*). Languages meet the first prerequisite by differing considerably in their color lexicons (Berlin & Kay, 1969). And physics provides the objective measure of a physical dimension (wavelength and intensity differences in radiant energy within the range detectable by the human visual system) against which human psychophysical performance (color discrimination and categorization) can be assessed. The question can now be posed: do human psychophysical functions in this dimension vary as a function of the language-specific labelings? That is, do speakers of different languages perceive hue at all differently? Various measures for such a distinction have been taken, for example discrimination across hue boundaries (speed, accuracy, and confusability), memory, and comparison of normal vs. defective ("color blind") similarity spaces. By and large the results of such investigations have been interpreted as favoring the independence of perceptual and linguistic categorization (see, for example, Brown & Lenneberg, 1954; Heider & Oliver, 1972; Jameson & Hurvich, 1978). However, recent work from Kay and Kempton (1984) suggests that even for this apparently well-defined perceptual domain, labeling (categorical perception) effects are or are not obtained depending on very delicate choices as to experimental procedures and particular stimulus characteristics (see also Davidoff, Davies, & Roberson, 1999).

Even were such issues resolved, studies of color organization do not constitute a fair test of the Sapir–Whorf hypothesis on several grounds. For one thing, the last place we might expect malleability of human thought is in domains involving sensory representations shared with many non-human species (for discussion, see Lucy, 1992a). Moreover, hue does not seem to be so important or central a factor in the life of a culture, excepting maybe the subcultures of art classes and Montessori preschools.

Perhaps more promising as domains within which language might interestingly influence thought are higher-level cognitive representations and processes, for instance, the linguistic encoding of time, or of object and substance, where linguistic variation is apparent and has by some authors been interpreted as straightforward reflections of

underlying differences in thought. Whorf himself didn't provide any systematic evidence that these features really legislated the thought of their users, and this is no surprise. A severe difficulty in investigating how language interfaces thought at these more "significant" and "abstract" levels has been their intractability to assessment. As so often, the deeper and more culturally resonant the cognitive or social function, the harder it is to capture it with the measurement and categorization tools available to psychologists.[1]

1.2. Space as test-bed

During the past several years, a certain consensus has been reached on a domain that may fall suitably between the two extremes of testability and ineffability; namely, spatial location and motion. This is a semantic dimension that (like hue perception) is fundamental to humans as well as all navigating creatures. Even so, the encoding of motion and space varies significantly across languages in ways that can be put under the psychologist's microscope. Much recent work in lexical semantics and psycholinguistics has been devoted to just this issue (see, for example, Bloom, Pederson, Nadel, & Garrett, 1996; Choi & Bowerman, 1991; Cummins, 1998; Gentner & Boroditsky, 2001; Jackendoff, 1990; Levin, 1985; Levin & Rappaport Hovav, 1992; Levinson, 1996; Li & Gleitman, in press; Munnich, Landau, & Dosher, 2001; Naigles & Terrazas, 1998; Pederson et al., 1998; Slobin, 1991, 1996a; Talmy, 1975, 1985, 2000; and Section 6 below).

Some of these studies of spatial encoding investigate relations that obtain between static objects, relations which languages encode quite variously. For example, even a single language may express the same or closely similar spatial relation by several lexical items: compare *Mary is sitting to the right of/next to/near Jim*. Others pertain to the linguistic description of dynamically unfolding displacement of objects and agents in the observed environment, including the causes, intentions, and goals of their movement. As such, these latter seem to us particularly promising testing ground for investigating how language may influence our conceptualizations of the world. Not simply lexical options, the issues here concern linguistic machinery for how entities and their movements and relations are expressed as predicate–argument structures; that is, how motion is syntactically realized.

Accordingly, the experiments in Sections 4 and 5 of this paper concern linguistic variability in the representation and expression of motion events. The remainder of this

[1] Returning for a moment to the topic of hue perception, Lucy (1992a) has objected not so much to this perceptual domain itself as a test-bed for linguistic determinism, but rather that psychologists have studied hue terminology independent of its cultural use and significance; he questions the "theoretical vision of a decontextualized 'natural' word-object relationship" (p. 185). For instance, in some communities *purple* is associated with royalty; in others, with rage; in yet others, with prose (perhaps Lucy would not subscribe to these particular examples, but they seem in accord with the gist of his discussion). If this objection holds, then color terminology is the worst of both worlds for learning how language might and might not affect thought – there is no shared conception any more than there is a shared terminology, so one can't hold one of these invariant (or manipulate it) so as to assess its effect on the other. Notice, though, that the more correct Lucy is in adopting this stance the more difficult in general it will be to compare the causal relations amongst three co-varying aspects of the human condition: language, social context, and concept. Indeed, much anthropological discussion is to the effect that parsing up the world to study such pieces of it inevitably disguises and distorts that which is to be studied. At the extreme, this stance threatens all positions with immunity.

introduction lays out such a variant linguistic property with particular emphasis on the instances of English and Modern Greek (Section 2), and then returns to the question of how this might bear on the relativist/universalist controversy (Section 3).

1.3. How languages differ: "typicality" vs. "grammaticality"

Before turning to the particulars of form-to-meaning mappings in the motion domain, it is of importance to lay out exactly what we (in common with most commentators in the present literature) count as a "linguistic difference" of the kind that might in principle bear on cognition. It is difficult if not impossible to accuse some language of being altogether lacking in the formal resources to express an idea that is codable in some other language. To be sure, languages differ in their lexicons in ways that seem intimately related to cultural preoccupations. The rich characterological terminology of diaspora Yiddish and the metaphors of falconry in Elizabethan English are frequently cited examples. But in the end one can translate the semantics of such words into other languages through phraseology that, though it may lose some of the color and natural force of the original, yet conveys the semantics passing well. This is similar for syntax-to-semantics mapping distinctions such as the count–mass distinction[2] or the Path–Manner distinction that forms the focus of the present paper. Neither Whorf nor Sapir, nor the commentators now discussing linguistic relativity in the current literature, have to our knowledge any investment in denying the pervasive encoding equivalences among languages in this sense. What is at issue, rather, are differences in *linguistic practice* and the cognitive effects these may engender.

Whorf was explicit in holding that the cognitive distinctions he was interested in would be a function of variation in "fashions of speaking" or typical linguistic practice, as opposed to categorical differences between invariances. The modern literature is explicit on this point. For instance, Pederson et al. (1998) writing on cross-linguistic differences in spatial encoding take it as a core feature of their investigations that "we looked at patterns of language use, not just grammatical descriptions" (p. 559) and "instead of comparing grammar alone, we are comparing linguistic PRACTICE – the meaning patterns that consistently emerge from domain-directed interactive discourse of members of a given speech community." (*ibid*, p. 565). (The change in theoretical fashion is in fact well highlighted by the fact that the disclaimer "just a statistical difference" in the not-so-distant past literature is now replaced by "just a grammatical difference".)

However, quoting recent literature may be an insufficient way to justify the experimental foray to which we next turn. A skeptic may still hold that statistical tendencies or "mere trends" in language use wouldn't be likely to have important nonlinguistic cognitive consequences and that any negative outcome of examining such probabilistic rather than categorical linguistic distinctions leaves the "real" Whorfian Hypothesis intact.

[2] We refer here to the fact that some languages, including English, have morphological means that formally mark the distinction between a mass (*water*) and a bounded object (*a dog*) whereas Japanese and many other languages do not (for discussion see Imai & Gentner, 1997; Lucy & Gaskins, 2001). Obviously, Japanese speakers nevertheless can distinguish between objects and the substance of which they are made, and their words for, say, 'wood' and 'chair' embody such a distinction. The question is only whether, all the same, there might be subtle cognitive effects of the linguistic machinery via which this distinction is made.

Such a stance would be hard to defend for two main reasons. We have already mentioned the first: if attention is to be concentrated only on cases where one language *can* while another language *absolutely cannot* express some idea, then there is no defensible phenomenon to investigate. The other reason is much more important. The past 20 years of perceptual, linguistic, and cognitive investigation in the information-processing framework have heavily documented the notion of contingent frequency as fundamental for describing both animal and human behavior (Gallistel & Gibbon, 2000; Rescorla & Wagner, 1972; Saffran, Newport, & Aslin, 1996). In particular, both language learning and on-line processes in language comprehension are realistically described only in terms of theories that are sensitive to the statistical distribution of relevant information (e.g. Gillette, Gleitman, Gleitman, & Lederer, 1999; Pearlmutter & MacDonald, 1995; Trueswell, 1996, *inter alia*). Thus, there is every reason to seek influences of language on thought by examining what speakers of a particular language community *do* do, apart from what under some other circumstances they *could* do if the design of their particular language was the only motive constraint.

2. The linguistic encoding of motion events

2.1. The Manner–Path distinction: cross-linguistic differences

Consider a simple motion scene: a man is running across the street. Human languages offer the means to parse this scene into a number of distinct encodable parts. For instance, language after language makes it possible to refer to the man separately from the ground he traverses (the street), to trace his trajectory (crossing), to comment on the details of his movement (running or hopping), to note whether the motion was externally caused or spontaneous (transitivity, *inter alia*), and so forth. The recurrence of certain meaning features in the linguistic description of motion has led to the supposition that the cognitive partitioning of motion events along certain dimensions is "natural", indeed universal (Landau & Jackendoff, 1993; Miller & Johnson-Laird, 1976; Talmy, 1985). On this view, widely accepted within cognitive science, linguistic representations of motion pick out a subset of the spatial–mechanical conceptual distinctions for which humans are cognitively prepared.

At the same time, along with this cross-linguistic identity of encoding there is systematic language-specific variance both in the commonly-used motion vocabularies of languages, and in the ways that languages conflate the elements of motion events inside grammatical structures. We will concentrate experimental attention on one such distinction in the linguistic encoding of motion events: *the Path–Manner distinction* (first discussed by Talmy, 1975). In what we will call **Manner languages** (e.g. English, German, Russian, Swedish, Chinese), manner of motion is typically encoded in the verb (e.g. *walk*, *run*), while path information appears in nonverbal elements such as prepositional phrases (*across the street*). In **Path languages** (e.g. Modern Greek, Spanish, Japanese, Turkish, Hindi), the verb usually encodes the direction of motion (e.g. *cross*, *ascend*), while the manner information is (optionally) encoded in gerunds or prepositional

phrases (and, in some cases, path verb–manner verb compounds).[3] Here are the typical renditions in the two languages we will investigate in the present paper:

(1) English

The man	walked	across	the street.
FIGURE	MOTION + MANNER	PATH	GROUND

(2) Modern Greek

O andras	dieshise	to dromo (me ta podia/perpatontas)
'the man	crossed	the street (on foot/walking)'
FIGURE	MOTION + PATH	GROUND (MANNER)

These are "typical" but not exclusionary ways of speaking (in Whorf's terminology, "fashions of speaking") within languages, as we discussed earlier (see Section 1.3). Manner languages do have common verbs which encode path information (e.g. English *enter*, *exit*, *ascend*, *descend*, *cross*, *turn*). And Path languages have ordinary verbs encoding manner of motion (e.g. Greek *kolibo* 'swim', *treho* 'run', *peto* 'fly', *horevo* 'dance'). However, the *preferred* lexicalization of motion events differs in the two language groups. English, for instance, has a rich collection of verbs which convey manner, but not directionality (*slide*, *roll*, *bounce*, *jump*, *stumble*, *limp*, *rush*, etc.); these verbs can be combined with a large set of adverbial or prepositional elements expressing path (*in*, *up to*, *across*, *out of*, etc.). By contrast, languages in the second group make sparser use of manner-of-motion verbs.[4] In Greek, the task of expressing change of location is mostly taken over by path verbs such as *aneveno* ('ascend'), *kateveno* ('descend'), *beno* ('enter'), *vgeno* ('exit'), which are frequently combined with prepositions such as *se* ('in/into'), *apo* ('from').

Furthermore, the use of manner-of-motion verbs in Path languages is subject to an interesting constraint (initially proposed for Spanish; see Aske, 1989; Jackendoff, 1990; Slobin & Hoiting, 1994): it seems that manner-of-motion verbs cannot be readily combined with path prepositional phrases to denote movement which involves some sort of bounded (completed, traversed) path. A typical instance of the restriction is given in (1): Greek in this case disallows the compact way of packaging manner and destination information that the English example captures. The most natural way of encoding this event in Greek is to omit manner information altogether (cf. (2)). Thus,

[3] Talmy (1985) used the terminology "satellite-framed" vs. "verb-framed", respectively, to describe this distinction. These terms again relate to where the "schematic core" of the motion event is placed syntactically: either in some modifying phrase or inside the verb itself. We adopt the Manner–Path language distinction in this paper because it offers a convenient shorthand for the meaning elements typically encoded in the main verb.

[4] Previous discussions have occasionally taken the stronger position that Greek-like languages have restricted motion verb vocabulary *in general* (the discrepancy with English-type languages being stronger in the case of manner-of-motion verbs). For instance, this is the conclusion reached by Sebastián and Slobin (1994) on the basis of a corpus count of attested verb types in English and Spanish narratives. But at the cost of belaboring a point we have emphasized throughout, restricted usage does not necessarily imply impoverished linguistic resources; in order to establish the latter, one would have to perform nothing less than a dictionary count. In fact, even the usage data are equivocal. Naigles, Eisenberg, Kako, Highter, and McGraw (1998), drawing on elicited descriptions of motion, conclude that there is no evidence that Spanish speakers use overall fewer motion-verb types than English speakers (see also Wienold, 1995).

the 'bounded path' constraint indirectly contributes to the more limited use of manner-of-motion verbs in Path languages.[5]

2.2. Navigating between Path and Manner languages

A first indication that the Path–Manner distinction "matters", in terms of how a scene is to be conceptualized and represented, is that translators notoriously run into trouble on this issue. Slobin (1996a) provides a very useful and psycholinguistically relevant analysis of such translation glitches. He showed (using the English–Spanish contrast in describing movement in space) that the greatest source of difficulty for Spanish translators was to retain an accurate version of the manner-of-motion descriptions of the English original. English translators could with lesser effort be more faithful to the Spanish text. Moreover, English translators were found to add manner information to the Spanish original, while Spanish translators often omitted manner information provided in the English text.

What goes for translators seems to go for ordinary users as well. Several recent studies show that speakers of Path and Manner languages characteristically describe different aspects of the same motion scenarios. These studies compare the linguistic behavior of adult speakers (usually of Spanish vs. English) in a variety of tasks involving the description of pictorially presented motion scenes. The findings reported by Slobin (1991), Berman and Slobin (1994), Naigles and Terrazas (1998), and Naigles, Eisenberg, Kako, Highter, and McGraw (1998) confirm that English speakers tend to describe motion with manner verbs, while Spanish speakers use predominantly path verbs. Similar preferences are attested in a corpus of children's narratives: after surveying this body of data, Sebastián and Slobin (1994, p. 262) conclude that, in Spanish, "manner is rarely attended to, at any age".

3. Mental organization and the language of space

The findings just cited interface only tangentially, if at all, with full-blooded Sapir–Whorf doctrines of linguistic determinism: the position (sometimes expressed by Slobin as "thinking for speaking") that learners (and adult speakers) mobilize language-specific categories so as to talk and understand is not controversial and is explicit in any information-processing approach to speech and comprehension, at all levels. After all, the very musculature of the vocal tract must be mobilized differently to express the same concepts, from language to language. One must round the lips in Greek (*arostos*) to speak of illness, but not so in English. Just so, as Slobin notes, the thoughts about the things and their comings and goings need to be organized in accord with the word-forms and the phrasal organizations of the particular tongue if its users are to speak at all.

However, it is possible to take a much stronger stand; namely, that such variations have a pervasive effect on nonlinguistic thought. It is often a short hop from noticing that

[5] The nature and application of the bounded path constraint are not properly understood. First, there is a small class of manner-of-motion verbs that do allow bounded path readings. We estimate that, in Greek, this corresponds to one-third of the set of manner-of-motion verbs. Second, even though several (and possibly all) Path languages appear to pose a similar constraint, synonyms across languages do not always behave identically with respect to the restriction. For some discussion, see Levin and Rapoport (1988), Levin and Rappaport Hovav (1995), Napoli (1992), and Folli and Ramchand (2001).

linguistic usage differs cross-linguistically to drawing nonlinguistic implications. Berman and Slobin (1994), commenting on possible consequences of the cross-language Path–Manner distinction, suggest that "…children's attention is heavily channeled in the direction of those semantic distinctions that are grammatically marked in the language" (p. 622). Similarly, in a study of English and Korean motion expressions, Choi and Bowerman (1991) suggest that the expression of conceptual elements (such as path) in the linguistic input may prompt or encourage children to pay special attention to them. For instance, Spanish or Greek children might especially notice paths upon encountering motion scenes; by contrast, the absence of clear and consistent linguistic marking of path might delay formation or deployment of the relevant conceptual distinctions in Manner languages.

3.1. An experimental prospectus: issues and predictions

There is certainly plausibility in the idea that differences in the linguistic encoding of motion events will have nonlinguistic effects on both learners and mature users. After all, even within a language the very same motion event is variably describable and perhaps concentrates the mind's eye on different aspects of it. For instance, *The mouse chased the elephant* and *The elephant fled (from) the mouse* describe the same action scenario and yet the two sentences bring the mouse and the elephant into differential focus (Fisher, Hall, Rakowitz, & Gleitman, 1994; Gleitman, 1990). What if, in this or a related regard, two languages disagreed *systematically and regularly* on how to frame events? Might this not train users to represent the world of locations and movements, actions and re-actions, doers and done-tos in contrasting ways? It certainly has seemed plausible to many recent commentators that such cross-linguistic encoding preferences would influence the construction of cognitive categories during, and as a consequence of, learning. Gentner and Boroditsky put it this way:

> Verbs and other relational terms – including those concerned with spatial relations – provide framing structures for the encoding of events and experience; hence a linguistic effect on these categories could reasonably be expected to have cognitive consequences. (Gentner & Boroditsky, 2001, p. 247)

But reasonableness is not sufficiency, as Gentner and Boroditsky also acknowledge. The claim that language influences nonlinguistic thought can never be decided by pointing to the linguistic facts alone. Granted that speakers of Path and Manner languages habitually *talk* differently. Granted therefore that their listeners would have to recover representations of the reference world from these variant language-specific encodings (as in Slobin's formulation). It still remains to ask whether the cognitive structures and representations achieved would vary too.

3.1.1. Recall and categorization in Greek vs. English

In the studies now presented, we will try to provide some evidence about whether such effects of language on nonlinguistic cognition actually exist in the motion domain. First, we ask whether speakers of English and Greek actually talk differently. We also ask whether any systematically differing preferences between English and Modern Greek in expressing paths and manners affect the way their speakers remember and classify motion

events. Furthermore, we investigate whether such nonlinguistic performance is affected by the degree of exposure to the target language. We do this by comparing the behavior of children and adult speakers of the two languages – who have been infected, so to speak, with their native tongues for varying lengths of time.

What particular effects (if any) should we expect the Path–Manner distinction to have on conceptual structure or processes? One prediction could focus on the lexical verb itself as the "informationally privileged" element, the expressive locus that will have memorial or categorization effects. After all, it is the lexical encoding preferences of the verb that provide the very name for the Path–Manner typology. Moreover, as we saw (example in (2)), very often the Path-language speaker will omit mention of manner altogether; therefore, manner descriptions are not constructed "for speaking" as regularly for these languages as for manner-verb languages. On either or both such grounds, we might expect that Path-language speakers must be specially sensitive to the path, and Manner-language speakers to the manner. This guess at a cognitive consequence has been the stance adopted by commentators we cited earlier: the fact that path-verb speakers often *omit mention of manner* might be interpreted to suggest they *don't as regularly attend to manner properties* of observed motion scenarios.

Just as reasonable, however, is the hypothesis that speakers of Manner languages will be more sensitive to path. According to such an argument, the path is exhibited independently ("foregrounded") on the surface of the motion verb sentence in manner-verb languages (*across the street*) rather than being hidden and wrapped up inside the meaning of the verb. Conversely, the manner is foregrounded in path-verb languages in gerunds, etc. (*me ta podia* – 'on foot'; for discussion of this perspective, see also Talmy, 1985). So long as there is a measureable relation between the language spoken and performance on cognitive tasks, in either direction, this can be interpreted as suggestive of a "Whorf-like" effect. Specifically, the experiments and analyses next reported are designed to examine four hypotheses:

- *Hypothesis 1: Greek and English speakers express path and manner differently in tasks that require them to describe a depicted motion scene*. This is simply a replication of studies and analyses we have mentioned earlier. It simply means that Greek speakers speak Greek and English speakers speak English. But this replication is required in order to show that, using our own materials and instructions, the two populations will diverge in their speech forms according to broad linguistic-typological styles.
- *Hypothesis 2: Memory and/or categorization performance for motion depictions will vary for speakers of the two languages*. This is the linguistic-relativistic prediction. On this account, differences between Manner and Path languages in the frequency and salience with which path vs. manner are encoded should result in systematic differences in how people in each language group attend to and process path vs. manner information in nonlinguistic cognitive tasks. In contrast, if we find that Greek and English speakers fall together in their memorial and categorization performance (while differing in their linguistic performance, Hypothesis 1), this is suggestive of the opposed universalist prediction.
- *Hypothesis 3: Because the language patterns (both the lexical items themselves, and their prototypical contexts of use) are learned, we expect to see Manner–Path expression to diverge more strongly in adults than in young children, within a language community*.

The idea here is that younger speakers may utter only a few, quite general, verbal items (perhaps *come* and *go*) with typological differences becoming manifest only as the stock of lexical items increases.

- *Hypothesis 4: Because the language patterns and their prototypical contexts of use are learned, nonlinguistic consequences (here, memory and categorization performance) will diverge progressively over age*. A similar position (although in a different context) can be found in the influential studies of John Lucy (Lucy, 1992b; Lucy & Gaskins, 2001) whose Tzeltal-speaking and English-speaking subjects behave similarly in cognitive tasks until about the age of 9, whereupon they diverge in ways that Lucy assigns in large part to growing expertise with the fine statistical structure of usage in the two languages.

3.1.2. Recall and categorization as a function of task upon task (encoding effects)

Each subject in these experiments performs two tasks. One is to verbally describe a motion scene. The other is either to remember that scene at a later time (Experiment 1) or to match it to another scene (i.e. to categorize it; Experiment 2). The order of the tasks varies in the two experiments. But in both cases, performance on the first task may well influence performance on the second. We assess such task-to-task influences in each experiment by means of an item analysis that collapses across language.

4. Experiment 1: Recognition memory

4.1. Participants

Participants were monolingual native speakers of either English or Modern Greek grouped into three age groups. The Young group included 38 English-speaking children between 4;2 and 6;0 years (mean age 5;3) and 38 Greek-speaking children between 4;4 and 7;2 years (mean age 6;2). The Middle group included 39 English-speaking children between 10;4 and 12;8 years (mean age 11;3) and 39 Greek-speaking children between 9;9 and 12;3 years (mean age 10;0). Finally, the Adult group included 20 English-speaking adults between 19;2 and 34;6 years (mean age 24;0) and 21 Greek-speaking adults between 18;1 and 50;8 years (mean age 29;7). Some of the English-speaking adults received course credit for participating.

4.2. Method

4.2.1. Materials

The stimuli for Experiment 1 consisted of a set of 8.5×11 inch black-and-white drawings adapted from the well-known Mayer (1969) frog stories for children.[6] For the purposes of this study, the original pictures were redrawn by an artist to simplify the scenes by removing possible distractors (Fig. 1, panel 1.1 and Fig. 2, panel 2.1). We

[6] The vivid illustrations of motion scenes in these stories have been used successfully in previous studies to elicit verbal descriptions of movement from both children and adults across a number of languages (Berman & Slobin, 1994).

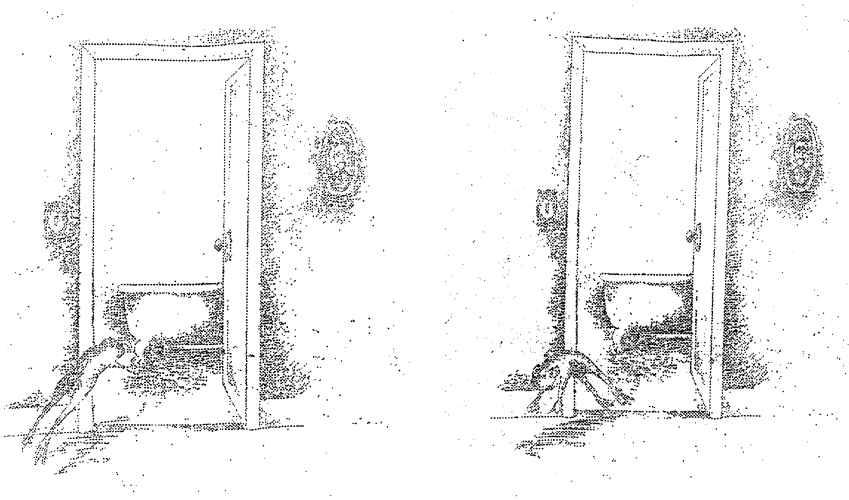

Fig. 1. Example stimuli for Experiment 1: path variation. (1.1) Session 1 item (frog jumping into a room). (1.2) Session 2 item (frog jumping out of a room).

also created variations of each scene by systematically altering either the path or the manner of the original movement. For example, to create a path variation, one of the original pictures showing a frog hopping into a room was altered to depict the frog hopping out of the room (Fig. 1, panel 1.2). To create a manner variation, a picture showing a boy jumping over a log was changed to show the boy tripping over the log (Fig. 2, panel 2.2).

The original stimulus set included six action scenes. Two of the action scenes failed to elicit motion descriptions from nearly all subjects, most likely because of the static single-picture presentation format. Moreover, subjects were not able to detect alterations to these two targets at a level that was above chance. Because these items did not yield informative data, we excluded them from all subsequent analyses. Of the remaining four scenes, one had a manner variation only, another had a path variation only, and two had both path and manner variants. The final set of four scenes with their respective variations is listed in Table 1.[7]

4.2.2. Procedure

Participants were tested individually by a single experimenter in two sessions 2 days apart. During Session 1, subjects were presented with the set of target pictures and asked to describe them. In Session 2, subjects were presented with a second set of pictures. The experimenter told the subjects that these pictures could be either the same as the ones they had seen in the previous session or different in some respect; participants were then asked to judge for each picture whether it was the same or different.

Each subject participated in four trials. Participants always viewed all four original action scenes in Session 1. In Session 2, for each of the scenes, that subject could view

[7] The two discarded items depicted a boy swinging from a rope and a boy climbing up a tree.

Fig. 2. Example stimuli for Experiment 1: manner variation. (2.1) Session 1 item (boy jumping over a log). (2.2) Session 2 item (boy stumbling over a log).

either the identical scene again or an altered scene, in which either the path or the manner had been changed. For clarity of exposition, we will refer to a particular combination of a Session 1 scene plus one of the possible corresponding Session 2 scenes as an item. Thus, for scenes in which both path and manner variations existed, three possible items could be constructed: the original scene plus an unaltered repetition; the original scene plus its path variant; and the original scene plus its manner variant. For scenes with just one altered variant, two possible items could be constructed – an unaltered item and an altered item. Any individual subject would receive only one of the possible item types for each scene. For scenes with two item types, approximately half of the participants saw each type, evenly distributed across language and age groups. For scenes that yielded three item types, approximately one-third of the participants saw each type, again evenly distributed across language and age groups. All together, there were ten unique items: four items with no changes, three with path variations, and three with manner variations.

4.2.3. Scoring

Any given trial can be classified as representing No Change from the original picture, a Path Change, or a Manner Change. Because of the way items were distributed across individual subjects, participants received different numbers of each type of trial. Three separate dependent measures were derived by calculating by participant the proportion of correct judgments made on the recognition task for Manner Changes, Path Changes and No Changes.

4.3. Results

4.3.1. Analysis of linguistic descriptions (main verb)

To see whether Greek and English speakers did, in fact, differ in their linguistic descrip-

Table 1
Stimuli for Experiment 1

Session 1	Session 2: path changes	Session 2: manner changes
1. A boy is jumping over a log.		A boy is stumbling over a log.
2. A frog is jumping into a room.	A frog is jumping out of the room.	
3a. A frog is hopping off a turtle.	A frog is hopping on a turtle.	
3b. A frog is hopping off a turtle.		A frog is falling off a turtle.
4a. A boy is diving off a cliff.	A boy is diving down a cliff.	
4b. A boy is diving off a cliff.		A boy is tumbling off a cliff.

tions of the pictures they viewed during the first session, we coded the main verb in each description as either Manner, Path, or Other. In order to perform an analysis of variance (ANOVA) we derived a dependent measure that was the total number of items on which a speaker used a manner verb as a main verb.[8] A two-way ANOVA was performed on this measure with Language (Greek vs. English) and Age Group (Young, Middle, Adult) as the independent factors. There was a main effect of Language ($F(1, 193) = 67.234$, $P = 0.0001$), but no effect of Age Group and no interaction. English speakers used a manner verb as the main verb on a mean of 3.60 out of 4 items, while Greek speakers used manner verbs on a mean of 2.86 items. Correspondingly, Greek speakers were much more likely to use path verbs as the main verb, as is shown in Table 2.[9]

4.3.2. Analyses of recognition task scores

Recall that on the recognition task, each trial could involve No Change from the drawing shown in Session 1, a Path Change, or a Manner Change. Because different subjects received different numbers of each kind of trial, we calculated a score summarizing the proportion correct for each subject for each kind of trial. A separate two-way ANOVA with Language (Greek vs. English) and Age Group (Young, Middle, Adult) as independent factors was performed on each of the three dependent scores for the recognition task (i.e. on the proportion of correct judgments on Manner Change trials, on Path Change trials, and on No Change trials). There was no main effect of Language on any of the three measures; nor did Language enter into a significant interaction with Age Group. Fig. 3 illustrates the absence of a Language effect on participants' ability to detect whether or not the path or manner was altered. There was a main effect of Age Group on the proportion of correct judgments on trials in which the Manner was altered ($F(2, 132) = 9.209$, $P = 0.0002$), with children in the Young age group scoring significantly lower (mean = 0.389) than either children in the Middle age group (mean = 0.710) or the

[8] The decision to count manner rather than path verb choice is arbitrary. Since there were relatively few verbs that could not be coded as either manner or path verbs (6.25% for the Greek speakers and 3.5% for the English speakers), the pattern for manner verbs is almost the inverse of the pattern for path verbs. An analysis using path verbs rather than manner verbs as the dependent measure yields analogous results with the same significance levels.

[9] It is worth noting that, in the English data, 68.3% of all manner verb uses co-occurred with path modifiers. The corresponding percentage for the Greek data is only 28.8%. The difference is to be expected given the typological differences described in Section 2 (especially the bounded path constraint).

Table 2
Verbal descriptions elicited in Experiment 1

	English (%)	Greek (%)
Path V	7.5	24
Manner V	89	69.75

Adult group (mean = 0.679) ($P < 0.05$, Games–Howell post-hoc test). There was no similar main effect of Age Group for trials involving Path Changes.

4.3.3. Collapsing across language: an item analysis for potential encoding effects

Each subject entered into the experimental setting just described by being confronted with a picture, which he or she visually inspected and, presumably, mentally represented. This initial encoding of a picture could have two interrelated effects: first, if the subject encoded the scene under a path representation this might cause him/her to describe it with a path verb (e.g. *The frog is entering the room*). And second, that very encoding and verbal description might survive over the 2 day interval and impact the memory performance. On this hypothesis, encoding performance predicts memory performance but is not a language-specific (that is, "Whorf-like") effect: it would occur in both languages, for sometimes an English speaker did utter a path verb and sometimes a Greek speaker uttered a manner verb. Such an effect can be extracted from our data set by an item analysis that collapses across language. The statistical question is this: suppose we index the subject's initial representation of the scene by his/her verbal description, i.e. if s/he describes a particular item with a path verb, will s/he be more likely to notice a path change alteration 2 days later?

To evaluate this question, we had to analyze for encoding effects taking into account the several ways in which the speaker could have expressed the path and manner elements in the depicted scenes. This meant going beyond the main verb to consider the total informa-

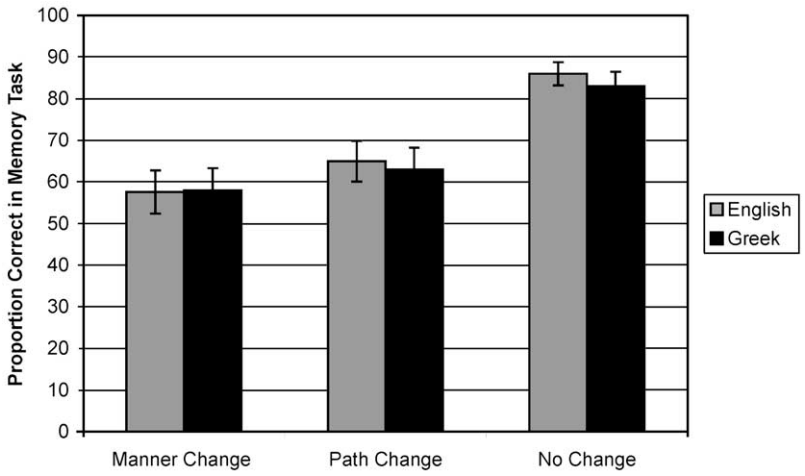

Fig. 3. Recognition task results in Experiment 1.

Table 3
Types of information in elicited descriptions in Experiment 1

Information in sample	Lexical-syntactic conflation patterns			
Path-Only	**Path V (bare):**			
	The frog	is entering.		
	O vatrahos	beni.		
	Path V + path modifier:			
	The frog	is going	into the room.	
	O vatrahos	pigeni	sto domatio.	
Manner-Only	**Manner V (bare):**			
	The frog	is jumping.		
	O vatrahos	pidai.		
	Manner V + manner modifier:			
	The frog	is jumping	briskly.	
	O vatrahos	pidai	apotoma.	
Combination	**Path V + manner modifier:**			
	The frog	is entering	jumping.	
	O vatrahos	beni	pidontas.	
	Path V + manner modifier + path modifier:			
	The frog	is going	into the room	jumping.
	O vatrahos	pigeni	sto domatio	pidontas.
	Manner V + path modifier:			
	The frog	is jumping	into the room.	
	O vatrahos	pidai	sto domatio.	
	Manner V + manner modifier + path modifier:			
	The frog	is jumping	briskly	into the room.
	O vatrahos	pidai	apotoma	sto domatio.
	Manner V + path V:			
	The frog	is entering (the room)	and jumping.	
	O vatrahos	beni (sto domatio)	ke pidai.	
Other	**Path Vs (stacked)**			
	The frog	is entering (into the room)	and advancing.	
	O vatrahos	beni (sto domatio)	ke prohorai.	
	Manner Vs (stacked)			
	He is stumbling	and falling (down).		
	Skontafti	ke pefti (kato).		
	Irrelevant Vs			
	The frog	is looking	into the room.	
	O vatrahos	kitai	sto domatio.	

tion provided by the speakers' responses. To do so we summarized the verbal descriptions elicited in this experiment in a coding scheme which included Path-Only, Manner-Only and Combination patterns. The Path-Only class included either bare path verbs or a path verb plus a path modifier. The Manner-Only class contained either bare manner verbs or a manner verb plus a manner modifier. The Combination class included a variety of cases in which the relevant aspects of the semantics of the verb differed from the type of information encoded by the modifier; moreover, it included multi-clausal responses which contained both a manner and a path verb (with or without additional modifiers). Table

3 gives a summary of the coding scheme with some examples from both languages (most of the examples were actually used as a description of the picture in Fig. 1.1).[10]

We now asked quantitatively whether there was a systematic relationship between the kind of information a speaker had included in his/her verbal description on a given item (Path-Only, Manner-Only, or Combination) and his/her memory performance on the same item, in case that item involved a Path or Manner Change in Session 2. A separate χ^2 analysis was performed on each of the six items that involved a change in the picture from Session 1 to Session 2. Each analysis took the form of a 2×3 contingency table for each item, crossing the kind of information that was included in the subject's verbal description (i.e. Path-Only, Manner-Only, or Combination) with whether the subject was correct or not in detecting the change that was made on that item.[11] For example, this set-up allows us to test whether subjects who included, say, path information in their verbal descriptions were more likely to detect a change to the path than subjects who did not encode path information in their descriptions. The χ^2 analyses revealed no significant effects: whether or not a subject had originally described an event in path or manner terms failed to be predictive of a difference in memory for path or manner.[12]

4.4. Summary

The speakers of Path and Manner languages differed in their likelihood of describing motion scenes with path vs. manner verbs (Hypothesis 1 was supported). The question was whether such differences in verbalization in Session 1 would predict memory performance during Session 2. We assessed this question in two ways, relevant to two interpretations of what such a correlation would mean. First, we asked whether speakers of the Path language would be more likely to utter path verbs and also more likely to remember path alterations, and similarly for the speakers of the Manner language. Such an effect would be a function of specific language (English vs. Greek) and independent of specific item consistency, a potentially Whorf-like effect. We found no such effect for either children or adults (Hypotheses 2 and 4 received no support). Second, we asked by means of an item analysis that collapsed across the two language groups whether the subject's original encoding of the scene (as indexed by his or her entire verbal description, i.e. choice of path or manner information, or a combination thereof) would influence memory for that scene 2 days later (as indexed by that subject's recognition of a path or manner alteration in the new picture). Such an effect would be independent of language, appraising the subject's within-item consistency. Again, no such effects were found.

[10] We excluded from this analysis descriptions with stacked path or manner verbs (since they may have included modifiers of a different semantic type than the verb) and irrelevant (non-motion) verbs (the combined proportion of these two constructions was only 2.5% for the Greek data and 5.4% for the English data).

[11] As we would predict from the typology of path and manner verbs, the occurrence of path verbs with manner modifiers is very rare. The majority of instances coded in the combination class are manner verbs with path modifiers. Because the distribution is so uneven, there is not a sufficient number of cases in the path verb subcategory to support a further numerical comparison.

[12] Statistics for the χ^2 analyses were as follows: item 1, $\chi^2 = 1.581$, d.f. $= 2$, $P = 0.4537$; item 2, $\chi^2 = 2.514$, d.f. $= 2$, $P = 0.2845$; item 3a, $\chi^2 = 0.743$, d.f. $= 2$, $P = 0.6897$; item 3b, $\chi^2 = 4.451$, d.f. $= 2$, $P = 0.1080$; item 4a, $\chi^2 = 1.542$, d.f. $= 2$, $P = 0.4625$; item 4b, $\chi^2 = 3.062$, d.f. $= 2$, $P = 0.0783$.

5. Experiment 2: Categorization

Several factors of design and measurement restrict interpretation of the results just presented. The picture-set was small, and it is possible that the static pictorial format made the recovery of path information somewhat harder than that of manner information.[13] For these and related reasons, differences between languages in the verbal description of motion were not as pronounced in this experiment as patterns reported more generally in the literature. Specifically, both groups seemed to focus on manner more than any other dimension of movement, even though English speakers were more likely to express this information in the main verb of their utterance than Greek speakers (see Naigles & Terrazas, 1998, for a similar finding, and discussion). Third, although participants generally scored above chance on the recognition task, the error rate was high (see Fig. 3). These limitations required further experimentation to resolve.

We therefore carried out a new experiment to investigate how the Path–Manner distinction enters into subjects' mental representation of motion events. This time motion events were represented in the more dynamic format of sequences of pictures and the task was categorization rather than memory.

5.1. Participants

Participants were monolingual native speakers of English and Greek who fell into two age groups. One group comprised 22 Greek-speaking 8-year-olds (range 7;2–9;2 years; mean 8;4) and 14 English-speaking 8-year-olds (range 7;5–10;0 years; mean 8;11). The second group comprised 21 Greek-speaking adults between 18;1 and 50;8 (mean 29;7) years of age and 20 English-speaking adults between 19;2 and 34;6 years of age (mean 24;0). These were the same subjects who participated in Experiment 1. They completed the categorization experiment immediately after Session 2 of the memory experiment (and before the debriefing session).

5.2. Method

5.2.1. Materials

Materials consisted of a picture-book containing eight sets of motion scenes. Each set consisted of three motion scenes in a match-to-sample format with one sample and two choices. The sample scene was presented on the left-hand page of the picture-book and the two choice scenes were presented on the opposite page. One of the choices preserved the path given in the sample while changing the manner of motion (a Manner Change); the other preserved the manner of the sample while changing the path (a Path Change). For instance, one of the sample scenes depicted a man running up the stairs. The path variation

[13] The fact that it is hard to identify paths from single static depictions of motion was probably the reason two of our initial items caused trouble for our subjects (and ultimately had to be discarded). As we will show in Experiment 2, sequences of pictures can solve this problem. For instance, in the present design of Experiment 1, subjects would have been able to infer the direction of motion (whether the boy was climbing up vs. down the tree) had there been a series of pictures depicting the progress of the motion.

Fig. 4. Example stimuli for Experiment 2. (4.1) Sample item (man running up the stairs). (4.2) Path variant (man running down a hallway). (4.3) Manner variant (man walking up the stairs).

for this case showed the same man *running down a hall*, while the manner variation showed the same man *walking up the stairs* (see Fig. 4).

To make dynamic information about both path and manner more accessible, we used a sequence of three digital color photographs (rather than a single drawing) to depict each motion event (this was done for both the sample and the two comparison events). All the events within each set involved the same animate protagonist (a man or a dog), or the same moving object (an airplane); an effort was made to keep the surroundings in all scenes as uniform as possible. A full list of stimuli is given in Table 4.

The eight items were arranged in two randomized orders, the second being the inverse of the first. These orders were counterbalanced across subjects within Language and Age groups. The presentation position of path vs. manner changes (top or bottom of the page) was counterbalanced across items in one fixed order.

A practice item depicted a non-motion event showing a man reading a book on a couch. In one of the choices, the man was shown sitting on the same couch drinking water, while in the other choice he was sitting on the stairs reading the same book. The practice trial was intended to emphasize that the matching decision should not be made on the basis of general similarity to the sample, but on the basis of similarity of action. This was important especially for our younger subjects.

Table 4
Stimuli for Experiment 2

Sample items	Manner changes	Path changes
1. stumble into a room	walk into a room	stumble down the stairs
2. jump into a room	walk into a room	jump off a chair
3. walk down the stairs	slide down the stairs	walk to the bookcase
4. run up the stairs	walk up the stairs	run down the hall
5. drive through a barn	walk through a barn	drive past a barn
6. sneak out of a room	walk out of a room	sneak into a room
7. jump off the stairs	fall off the stairs	jump on the couch
8. fly over the barn	fly upside down over the barn	fly around the barn

5.2.2. Procedure

Participants were tested individually by a single experimenter. Subjects were first presented with the practice trial and were asked to select the choice in which the man was "doing the same thing" as in the sample. Unlike the main phase of the experiment, the practice trial has a correct answer; if a subject chose incorrectly, the experimenter pointed out the mistake and gave appropriate feedback.

The interview then proceeded with the test trials with the same instruction on each trial (to select the choice in which the agent was doing the same thing as in the sample). After all trials were completed, participants were asked to describe each scene verbally. Even though no restrictions were placed as to the appropriateness or length of description, most subjects gave a one- or two-utterance response.[14]

5.3. Results

5.3.1. Analysis of linguistic descriptions (main verb)

The main verb in subjects' descriptions of the pictures was coded as Manner, Path, or Other. (If a response included two main verbs from different classes, this response was also coded as "Other".) A dependent variable was created by summing the total number of items (out of eight) for which the main verb was coded as Manner. This variable was then used in a two-way ANOVA with Language Group (Greek vs. English) and Age Group (Children vs. Adults) as the independent variables. The analysis revealed a main effect of Language Group ($F(1, 73) = 217.536$, $P = 0.0001$), with English speakers using manner verbs on an average of 5.029 items compared to 1.419 items for Greek speakers. There was also a main effect of Age Group ($F(1, 73) = 13.011$, $P = 0.0006$). Adults used more manner verbs in describing the sample photographs (mean = 3.585) than children did (mean = 2.361). The Language Group by Age Group interaction yielded a significant effect ($F(1, 73) = 17.541$, $P = 0.0001$). As illustrated in Table 5, the interaction was due to the fact that English adults and children differed from each other, with the English adults using more manner verbs, while Greek adults and children used manner verbs with

[14] Delaying the verbal description until after the categorization performance was so as to avoid contaminating the latter by any mechanism that would favor consistency (e.g. a recency or priming effect).

Table 5
Verbal descriptions for sample elicited in Experiment 2

Group	Mean manner verbs (out of eight sample descriptions)	Mean path verbs (out of eight sample descriptions)
Greek children	1.500	5.000
Greek adults	1.333	5.429
English children	3.714	3.643
English adults	5.950	1.550

similar frequency. As the table also shows, English children used fewer manner verbs than English adults but more manner verbs than either Greek children or Greek adults.

A corresponding analysis was done using the total number of path verbs produced by each participant as the dependent variable. Because of the low percentage of responses coded as "Other", the path data are essentially the complement of the manner data. The same ANOVA using the path verb dependent variable yields the same pattern of results with similar probability values (see also Table 5).

An age-related effect was obtained only for the English speakers, with the English-speaking children showing a less differentiated pattern, using manner and path verbs about equally (e.g. *jump*, *run*, and *come*, *go*, respectively). In Greek children, there is already a predominance of path verbs.

5.3.2. Analysis of categorization preferences

To analyze categorization preferences, we created a dependent variable consisting of the total number of trials on which participants chose the manner change as the best match to the sample.[15] A two-way ANOVA performed on this dependent variable with Language and Age Group as the independent factors yielded no main effects and no interaction between Language and Age Groups. Both English and Greek speakers, whether children or adults, chose manner variations on approximately half of the trials and path variations on the other half. As Fig. 5 illustrates, while English and Greek speakers show a pronounced difference in the types of verbs they use to encode motion events, the two groups are indistinguishable in their categorization preferences.

5.3.3. Collapsing across language: an item analysis for potential encoding effects

Recall that in Experiment 1 we asked (by collapsing by verbal description, across language) whether the subject's original verbal description, regardless of her native language, was related to subsequent memorial performance. That is, would having uttered, say, a Path-Only description to describe the event increase the likelihood that this subject, for this item, would show enhanced sensitivity to alteration in its path? Here, where the categorization task precedes the verbal description, we can again ask the question whether there are task-on-task effects: did the way the subject categorized the sample item (as indexed by categorization choice) predict subsequent verbal description? To answer this,

[15] Because the task was a forced choice between two alternatives, one could equally well use the number of trials on which the path variant was chosen as the dependent variable without affecting the pattern of the results.

we performed χ^2 analyses separately on each item, crossing the overall verbal description (Path-Only, Manner-Only, or Combination) offered by a subject for the sample item with his or her categorization choice (manner or path) for the same item. These analyses collapsed responses across Language and Age groups. Rather to our surprise, these analyses yielded no reliable support for the hypothesis that the subject's initial encoding of an individual scene (indexed by the categorization performance) influenced the later verbal description. For six of the eight items there was no significant relationship of any kind.[16]

5.3.4. Can we torture the data?

So far, our analyses have indicated that classification choices were correlated neither with stable typological patterns, nor with item-by-item preferences in linguistic encoding. We now turn to a final, even more specific analysis. Suppose that a speaker produced the *identical* verb in her description of both the sample and the same-path variant, but not for the same-manner variant. Would that speaker turn out, on average, to have also – for this item – chosen the same-path variant in the categorization task?

To find out, we tested whether manner verb matches (an identical verb uttered by the subject to describe the sample and its manner-matched variant) were more closely associated with manner categorization preferences and path verb matches were more closely associated with path categorization preferences.[17] This was done in a series of item-by-item χ^2 analyses relating type of verb match (Path Verb Match, Manner Verb Match, or No Match) to participants' categorization preferences on each item. These analyses were performed separately for Greek and English speakers. In neither language group did verb matches correlate consistently with categorization performance.[18]

5.4. Summary

We assessed the relation between categorization of the stimuli and cross-language verbal descriptions of them using several measures of the latter. The first finding was that Greek and English speakers differed in their tendency to express manner information (Hypothesis 1 was again confirmed). Moreover, this language difference grew over age

[16] Of the two items where there was such a significant relationship, item 3 ($\chi^2 = 5.230$, d.f. $= 1$, $P = 0.0222$) was consistent with the hypothesis (Path-Only information was more likely to co-occur with path choices and there were no Manner-Only descriptions); item 1 ($\chi^2 = 8.585$, d.f. $= 2$, $P = 0.0137$) contradicted it (both Path- and Manner-Only descriptions were more likely to coincide with path categorization preferences). Note that a more conservative P value adjusted for the number of analyses in this set would be less than 0.00625.

[17] In verb matches (a) the sample and only one of the variants shared the same verb, (b) neither the sample nor the variant contained any other verbs, and (c) other clausal components (modifiers, etc.) of the sample and the variant were ignored. We performed a separate analysis for modifier matches. It turns out that, even when a speaker used the exact same modifier to encode the sample and only one of the variants, they were not, in general, more likely to match the sample to that variant in the categorization task.

[18] For the Greek speakers, there were no significant relationships between match types and categorization preferences on any of the items. For English speakers, verb matches were related to categorization preferences only on items 1 and 3. For item 1 ($\chi^2 = 12.424$, d.f. $= 2$, $P = 0.0020$), matches on manner verbs were associated with manner categorizations. For item 3 ($\chi^2 = 6.170$, d.f. $= 2$, $P = 0.0457$), both path and manner matches predicted their respective categorization choices.

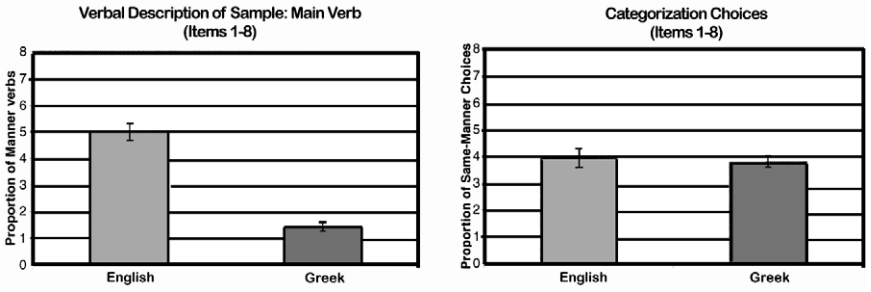

Fig. 5. Results from Experiment 2. (5.1) Linguistic task. (5.2) Non-linguistic task.

(Hypothesis 3 was confirmed). However, categorization of the visual stimuli did not differ across language or across age groups (Hypotheses 2 and 4 received no support). Item analyses that collapsed across language also showed no relation between categorization performance and subsequent verbal description. This nonrelation between the two tasks held true even where the subjects' verb label was identical for both the sample and one variant.

6. General discussion

The experiments just reported examined language–thought interactions in a particular domain of language difference: the Path–Manner distinction, famously analyzed by Talmy (1985). In two important ways, this distinction is of the kind that held particular interest for Whorf, and that he tried to relate to significant cultural patternings: first, the linguistic distinction is not (or not solely) some isolated lexical choice or single constructional type. Rather, as we saw, the Path–Manner distinction is played out in a nexus of linguistic selections (a "cryptotype") that are sometimes lexical, sometimes syntactic, and sometimes cross-clausal in their manifestations, and which map onto complex aspects of events. Second, and closely related, Whorf had in mind not the idea that some language can express a thought where some other language actually has a gap in expressive resources. Rather linguistic communities manifest probabilistic differences in "fashions of speaking" that he suggested would foster and reflect deep distinctions in the discourse of a culture.

Our experimental review was designed to further document the Path–Manner distinction, using Greek and English as the test languages, and then to ask whether the obtained language differences were related to performance in nonlinguistic cognitive domains. Hence, we tested language usage against categorization and memory performance. Here we first review the findings and the generalizations they suggest about the relative immunity of these cognitive functions to influence by language-specific semantico-syntactic design features. Thereafter we briefly take up the recent literature that, taken as a whole, documents clear performance differences as a function of particular language. We will try to say why the findings from these experiments and those reported herein are not really in conflict, despite some interpretations to the contrary.

6.1. Dissociations of linguistic and conceptual performance on the Path–Manner dimension

6.1.1. Do English and Greek differ in the required ways?

Experiments 1 and 2 redocumented the Path–Manner typological distinction between Greek and English. The verbal reports of subjects in both Experiment 1 (Table 2) and especially in Experiment 2 (Table 5) – where the stimuli were more dynamic and sampled more widely in the relevant domain – revealed language-specific distinctions of the predicted kinds. These differences grew with the age of the speakers tested. As we showed (Table 5), these age-dependent findings reside primarily in the English-speaking population. The verbal style of the Greek subjects did not change significantly over age in the relevant regards. This difference can be assigned to the necessity for English speakers to acquire a very large and in some ways unlimited vocabulary of manner-verb distinctions, i.e. in English one can *jet to Paris*, or *seven-forty-seven to China*.[19] It takes a significant time period within early and middle childhood to acquire this large vocabulary and its generative possibilities. Therefore, our youngest English-speaking subjects' verbal descriptions tended to stick with a small number of frequent verbs of motion, several of which (*come*, *go*) are path verbs, and thus look less different from their Greek peers than was the case for the older subjects.

6.1.2. Influences of language on nonlinguistic cognitive functions

We next turned to the question of how these clear linguistic distinctions map onto nonlinguistic thought. From Whorf's perspective there are causal links here and, though the influences go both ways, in this partnership "the nature of the language is the factor that limits free plasticity and rigidifies channels of development in the more autocratic way":

> …language is not merely a reproducing instrument for voicing ideas but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock in trade. Formulation of ideas is not an independent process…but is part of a particular grammar, and differs, from slightly to greatly, between different grammars. (Whorf, 1956, pp. 212–213).

To the extent that this is so, it is reasonable to expect that memorial performance for paths and manner of motion and categorization performance for paths and manner should differ systematically between the Greek- and English-speaking populations. Yet neither in the memory task of Experiment 1 nor in the categorization task of Experiment 2 did the English-speaking and Greek-speaking subjects differ as they should have if the latter population was essentially blind to manner properties of events.

One way to dismiss these findings is to treat them as mere null effects: difference probed for, but not found. The measures might have been too weak, the instructions might have been misleading, and so forth. From this perspective, to say that the experimental manip-

[19] True also to Whorf's complex vision, these manner innovations are fed by a cooperating set of language-particular options, including the license to shift between noun and verb without added morphology, and to chain nominals into compounds.

ulations provided no support for a postulated language-driven difference in conceptualization is not to say much. But such an interpretation of the findings ignores the fact that there are stable and strong outcomes in both experiments, along the relevant dimensions: *Greek speakers just like English speakers did attend to manner, and to the same degree, both to remember past depictions of motion and to categorize new ones*. Thus, demonstrably the lexical patterning of the specific languages did not bleed into subjects' performance in tasks that do not call on the linguistic categories specifically.

Our findings add to very recent experimental results (some of which were independently obtained at roughly the same time as our studies), which suggest that spatial cognition may not be a promising source of Whorfian effects. Munnich et al. (2001) compared English, Japanese and Korean speakers' naming of spatial locations and their spatial memory for the same set of locations. They found that, even in aspects where languages diverged (such as encoding contact/support spatial relations), there was no corresponding difference in memory performance among the language groups. In another cross-linguistic study, Gennari, Sloman, Malt, and Fitch (in press) found that, even though Spanish- and English-speaking adults provided different verbal descriptions for motion events, these subjects did not differ in what they remembered about these motion episodes. Related findings extend beyond the spatial domain: Malt, Sloman, Gennari, Shi, and Wang (1999) report that cross-linguistic differences in naming objects do not affect object categorization for Spanish-, English- and Chinese-speaking subjects.[20]

6.1.3. Language is no faithful mirror of our thoughts

Nothing is less surprising than that spatial organization should be largely immune to language influences. After all, like all navigating creatures from ground wasps to ballet dancers, Greeks and Americans must be attending to manner and path in much the same fashion lest they wander drunkenly off the edge of the earth. In the language and thought literature to which we have here attempted to contribute, language-sensitive effects are expected (by all parties) to be restricted to cognitive performances at a level far removed from the perceptual foundations of spatial acts and inferences. Yet even at these higher-order cognitive levels, we did not see changes in the salience of spatial dimensions as a consequence of how these are semantically encoded by the two languages. What is *really* surprising and potentially informative is that there were no measureable task-on-task effects in the item analyses for either experiment.

The idea motivating the item analyses is this: an observer visually inspecting a motion scene might represent it to himself or herself either as a path event or as a manner event. This could be true, on occasion, of speakers of both languages though the frequencies might differ across languages and would change even within a speaker from occasion to occasion. But whatever that initial representation, one might expect it to be reflected in memory performance in Experiment 1 or categorization performance in Experiment 2. The item analyses probed for just such effects, and did not find them. For the memory task,

[20] Billman, Swilley, and Krych (2000) have observed effects of linguistic labeling on recognition of motion events when labeling immediately precedes the nonlinguistic task (cf. also Gennari et al., 2002, for similar effects on categorization). In these cases, it is reasonable to assume that the 'nonlinguistic' task is actually solved by verbal mediation.

the analysis takes the subject's initial verbal description as indexing his or her initial representation of the scene, either as manner ("He's running") or path ("He's crossing the room"). If the verbal description actually indexes the subjects' conceptual encoding of the scene, it ought to predict memory performance. But the verbal label did not turn out to predict the subject's memory for path vs. manner alterations when this was tested 2 days later (Fig. 3). Similarly in Experiment 2, the analysis takes the subject's categorization (manner vs. path choice) as indexing his or her initial conceptual representation of the depicted scene. Again, this nonverbal performance failed to predict the subsequent verbal labeling of the scene (Fig. 5).

The fault clearly lies in the presuppositions of these analyses; namely, that verbal descriptions are tightly tied to (and therefore can serve as indices of) how the subjects have conceptually represented the materials. There is every reason to suppose that language is a far cruder instrument than this. People verbally describe the world to each other (and, sometimes, to themselves) using whatever formal and substantive resources the particular language makes easily available. What the item analyses reflect is that those verbal reports did not come anywhere near exhausting the observers' mental representation of the motion events that they were shown. Language use is in this sense sketchy (for a more general discussion of sketchiness, see also Fisher & Gleitman, in press). A Greek speaker's everyday and easily accessible vocabulary of motion verbs will cause him/her to call up a path verb to describe events most of the time, including those events that, for English speakers, are most readily nameable with manner verbs ("run" is more frequent than "ascend" or even "go up"). To paraphrase the Mad Hatter in this regard, while people generally mean what they say, it does not follow that they say what they mean. One tends to say whatever the language forms make it easy to express, leaving it to the listener to fill in the gaps.

Summarizing, the noneffects of the item analyses provide one more suggestive piece of evidence for the view that while any particular language is a partial vehicle for representing thought, its limitations and exactitudes do not impose themselves on the representation of experience. We do not literally "think in English" (or Greek, etc.). This perspective is in direct opposition to Whorf's strongest interpretations of language as "the program and guide" for mental activity. Rather, language seems to be "merely a reproducing instrument for voicing ideas" and not such a very faithful one at that. For most communicative purposes it is enough that speech evoke ideas in listeners, not that it literally render those ideas.

6.2. *Influences of the linguistic packaging on task performance*

The past decade's investigations have revealed cognitive influences of language on thought of quite another kind than we have been discussing so far. Languages package the same conceptual information in rather different ways that their users come to know in the course of learning (Bowerman, 1996; Slobin, 1991, 1996b). Talmy's *conflation patterns* for event representation, experimentally studied in the present paper, are a specific instance. As we discussed in Section 1 of this paper, both Greek and English represent motion events as comprised of the same proper parts – the figure (or the moving object), the path, the manner, the goal. In this sense, the conceptual representation of motion events

does not vary across the two languages. However, the parts themselves are distributed differently, in linguistic practice – that is, in the highly frequent verb vocabulary of users.

These implicitly known packaging differences have been shown to affect behavior in important ways. *When subjects are told* (typically, by the implicit means of introducing a nonsense word) *that the task is a linguistic one, semantically conditioned effects are usually observed*. In a beautiful example, Naigles and Terrazas (1998) showed their English- and Spanish-speaking subjects video scenes in which a person is depicted as moving toward some goal in some manner (say, a girl skipping toward a tree) accompanied by audio containing a nonsense verb in a path-biasing structure ("She's kradding the tree", i.e. perhaps, *approaching*, or its Spanish equivalent "Ella está mecando al árbol") or a manner-biasing structure ("She's kradding toward the tree", i.e. perhaps, *skipping*, or its Spanish equivalent "Ella está mecando hacia el árbol"). Both groups were guided by the syntactic structure, giving more judgments in favor of the path meaning for transitive verbs and more in favor of the manner meaning for intransitive verbs; that is, their conjectures were sensitive to the semantic implications of the syntactic structure (a procedure known as "syntactic bootstrapping"; Fisher, 1996; Fisher, Gleitman, & Gleitman, 1991; Gleitman, 1990; Naigles, 1990). However, in addition there was a clear influence of the lexical patternings that differed across the two languages, with Spanish speakers making far more path conjectures overall than English speakers. Clearly, then, speakers have mastered the statistics of syntax-to-semantics mappings in their languages and (at least in laboratory situations) will bring this knowledge to bear in deciding the meaning of such new verbs as *kradding* or *mecando*. In contrast, in Experiments 1 and 2 of the present paper, when the categorization or memory tasks are divorced (or "more divorced") from language itself, performance by speakers of different languages looks the same.

Effects related to those of Naigles and Terrazas come from the well-known findings of Imai and Gentner (1997). Following a design from Soja, Carey, and Spelke (1991) they asked American and Japanese children and adults to categorize novel stimuli – various oddly shaped objects made of novel materials (e.g. toothpaste stuck with sparkles). Responses were expected to differ to the extent that English (at least roughly) marks objects with a special morphology (*a horse*) but not substances (*water* rather than *a water*). Presented with a sample object/substance labeled neutrally ("This is my dax" or "Do you see this dax?"), subjects' extension patterns ("Can you show me another dax?") were investigated. A clear object bias was observed cross-linguistically. But for certain in-between or ambiguous test items (e.g. a lima-bean shaped lozenge made of semisoft wax) there was a clear language effect: English speakers favored the object over the substance interpretation for these cases while the Japanese speakers did not. Again we see the influence of language-specific packaging *when the task is a linguistic one*: the expression "this dax", though in principle neutral as between mass and count morphology ("this candle", "this wax"), is highly likely to be a count noun in English just because count nouns overwhelmingly outnumber mass nouns in English. But Japanese semantics provides no such statistical guidance. Where clear object/substance differences provide no or poor guidance, then, the subjects apparently invoked known language patterns. But as Mazuka and Friedman (2000) showed, the classifications made by speakers of these two languages fell together when the implicit language instruction was removed from the situation (see also Kay & Kempton, 1984).

Summarizing, several investigators have ingeniously shown that language users, even young children, have mastered the probabilistic form-meaning patterns of their languages. They know which kinds of concepts are likely to get bundled together within the word (or phrase). If a task is presented as a labeling one, and especially when the stimulus situation is ambiguous or inconsistent, the speakers are likely to invoke this knowledge, with the outcome that users of different languages will behave differently. But to the extent that language influence is experimentally lessened by the choice of stimulus materials or task instructions, as in the experiments we have presented, these differences among speakers of different languages tend to diminish or disappear too.

6.3. Final thoughts

We have tried to engage the question of linguistic relativity as it has resurfaced in the current psychological and anthropological literature, using motion representation and expression as the test case. Our findings suggest a good measure of independence between conceptual and linguistic representation. Such findings come as little surprise to recent proponents of the linguistic relativity position for they – as well as Whorf and Sapir themselves, except in their most enthusiastic moments – have no quarrel with the notion that certain immutable core concepts are shared across the species. Rather, these investigators have tried to show that in addition to these shared concepts, there are different ones too, literally caused and imposed by the language differences themselves. Sometimes this view is called "weak Whorfianism". Based on our findings, we have urged quite a different perspective: many apparent effects of language on thought are more appropriately interpreted as effects of language on language. In tasks that explicitly or implicitly call for knowledge of language, subjects can in fact draw upon that knowledge. Because the languages differ, speakers of the different languages will differ accordingly. But this proves no more (and no less!) than that English speakers speak English and Greek speakers speak Greek. The linguistic relativity question, interpreted nonvacuously, is whether by having learned these languages speakers differ in the very basis of their inductions, whether they are representing, categorizing, remembering, and reasoning in terms of a (partially) different set of experiential categories. In our view the answer to these questions tends to be no. Here we argued the case by pointing to the following generalization: the more language-like the subjects' task, the more speakers of different languages can be shown to vary in their performance; the more language is removed from the task situation, the more subjects exhibit their human conceptual commonalities.

Acknowledgements

We would like to thank Mark Baker, Liz Spelke, Jerry Fodor, Randy Gallistel, the two anonymous reviewers, and the members of the CHEESE seminar for comments and discussion. We also thank Jason Rosas for his help in carrying out Experiment 1 and George Pappas whose running up and down provided materials for Experiment 2. Our experiments took place at the Packer Collegiate Institute, the Old Pine Community Center and the Parent–Infant Center in Philadelphia, the Language School 'Ekpedeftiki' in Athens, the 3rd State Daycare and the 5th Elementary School in Vrilissia: we thank the

teachers and children for their cooperation. This research was supported by a Postdoctoral Fellowship to the first author from the Institute for Research in Cognitive Science, University of Pennsylvania, by NSF-STC Grant #SBR-89-20230, and by NIH Grant #1-R01-HD37507-01 to J. Trueswell and L. Gleitman.

References

Aske, J. (1989). Path predicates in English and Spanish: a closer look. In K. Hall, M. Meacham & R. Shapiro, *Proceedings of the fifteenth annual meeting of the Berkeley Linguistics Society* (pp. 1–14). Berkeley, CA: Berkeley Linguistics Society.

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

Berman, R. & Slobin, D. (1994). *Relating events in narrative: a cross-linguistic developmental study*. Hillsdale, NJ: Erlbaum.

Billman, D., Swilley, A., & Krych, M. (2000). Path and manner priming: verb production and event recognition. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 615–620). Mahwah, NJ: Erlbaum.

Bloom, P., Peterson, M., Nadel, L. & Garrett, M. (1996). *Language and space*. Cambridge, MA: MIT Press.

Bowerman, M. (1996). The origins of childrens spatial semantic categories: cognitive versus linguistic determinants. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–176). Cambridge: Cambridge University Press.

Brown, R., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, *49*, 454–462.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. *Cognition*, *41*, 83–122.

Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.

Cummins, S. (1998). Le mouvement directionnel dans une perspective d'analyse monosemique. *Langues et Linguistique*, *24*, 47–66.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398* (6724), 203–204.

Fisher, A., Hall, G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333–375.

Fisher, C. (1996). Structural limits on verb mapping: the role of alignment in children's interpretation of sentences. *Cognitive Psychology*, *3*, 41–81.

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

Fisher, C., & Gleitman, L. (in press). Language acquisition. In H. F. Pashler & C. R. Gallistel (Eds.), *Stevens' Handbook of Experimental Psychology. Vol. 1: Learning and Motivation* (3rd ed.). New York: Wiley.

Folli, R., & Ramchand, G. (2001). Getting results: motion constructions in Italian and Scottish Gaelic. In K. Megerdoomian & L. A. Bar-el (Eds.), *Proceedings of the 20th West Coast Conference on Formal Linguistics* (pp. 101–114). Somerville, MA: Cascadilla Press.

Gallistel, C. R., & Gibbon, J. (in press). Time, rate, and conditioning. *Psychological Review*, *107*, 289–344.

Gennari, S., Sloman, S., Malt, B., & Fitch, W (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge: Cambridge University Press.

Gillette, J., Gleitman, L. R., Gleitman, H., & Lederer, A. (1999). Human simulations of language acquisition. *Cognition*, *73*, 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

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

Imai, M., & Gentner, D. (1997). A crosslinguistic study of early word meaning: universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jameson, D., & Hurvich, L. (1978). Dichromatic color language: "reds" and "greens" don't look alike but their colors do. *Sensory Processes*, *2*, 146–155.

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

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217–255.

Levin, B. (1985). *English verb classes and alternations*. Chicago, IL: University of Chicago Press.

Levin, B., & Rapoport, T. (1988). Lexical subordination. In L. Macleod, G. Larson & D. Brentari (Eds.). *Papers from the Twenty Fourth Regional Meeting of the Chicago Linguistics Society* (pp. 275–289). Chicago, IL: University of Chicago Press.

Levin, B., & Rappaport Hovav, M. (1992). The lexical semantics of verbs of motion: the perspective from unaccusativity. In I. Roca (Ed.), *Thematic structure: its role in grammar* (pp. 247–269). Berlin: Foris.

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: at the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: crosslinguistic evidence. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (Eds.), *Language and space* (pp. 109–170). Cambridge, MA: MIT Press.

Li, P., & Gleitman, L. (in press). Turning the tables: language and spatial reasoning. *Cognition*.

Lucy, J. (1992). *Language diversity and thought: a reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. (1992). *Grammatical categories and cognition: a case-study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: a comparative approach. In S. Levinson & M. Bowerman (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge: Cambridge University Press.

Malt, B., Sloman, S., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Books.

Mazuka, R., & Friedman, R. (2000). Linguistic relativity in Japanese and English: is language the primary determinant in object classification? *Journal of East Asian Linguistics*, *9*, 353–377.

Miller, G., & Johnson-Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: a cross-linguistic comparison. *Cognition*, *81*, 171–207.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*, 357–374.

Naigles, L., Eisenberg, A., Kako, E., Highter, M., & McGraw, N. (1998). Speaking of motion: verb use in English and Spanish. *Language and Cognitive Processes*, *13*, 521–549.

Naigles, L., & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish: influences of language and syntax. *Psychological Science*, *9*, 363–369.

Napoli, D. J. (1992). Secondary resultative predicates in Italian. *Journal of Linguistics*, *28*, 53–90.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, *34*, 521–542.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language*, *74*, 557–589.

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

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Sapir, E. (1941). In L. Spier, *Language, culture and personality: essays in memory of Edward Sapir*. Menasha, WI: Memorial Publication Fund as quoted in Whorf (1956), p. 134.

Sebastián, E., & Slobin, D. I. (1994). Development of linguistic forms: Spanish. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: a crosslinguistic developmental study* (pp. 239–284). Hillsdale, NJ: Lawrence Erlbaum.

Slobin, D. (1991). Learning to think for speaking: native language, cognition, and rhetorical style. *Pragmatics*, *1*, 7–26.

Slobin, D. (1996). Two ways to travel: verbs of motion in English and Spanish. In M. Shibatani & S. Thompson (Eds.), *Grammatical constructions: their form and meaning* (pp. 195–219). Oxford: Oxford University Press.

Slobin, D. (1996). From 'thought and language' to 'thinking for speaking'. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). New York: Cambridge University Press.

Slobin, D., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: typological considerations. In S. Gahl, A. Dolbey & C. Johnson (Eds.), *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society* (pp. 487–505). Berkeley, CA: Berkeley Linguistics Society.

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. *Cognition*, *38*, 179–211.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics* (pp. 181–238), Vol. 4. New York: Academic Press.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 57–149). New York: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics* (Vols. 1 and 2). Cambridge, MA: MIT Press.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566–585.

Whorf, B. (1939). The relation of habitual thought and behavior to language. (Reprinted in *Language, thought and reality: selected writings of Benjamin Lee Whorf*, pp. 134–159, by J. B. Carroll, Ed., 1956, Cambridge, MA: MIT Press).

Whorf, B. (1941). Language, mind, and reality. (Reprinted in *Language, thought and reality: selected writings of Benjamin Lee Whorf*, pp. 246–270, by J. B. Carroll (Ed.), 1956, Cambridge, MA: MIT Press).

Whorf, B. (1956). In J. B. Carroll (Ed.), *Language, thought and reality: selected writings of Benjamin Lee Whorf.* Cambridge, MA: MIT Press. All page references to the 1995 edition.

Wienold, G. (1995). Lexical and conceptual structure in expressions for movement and space. In U. Egli, P. Pause, C. Schwarze, A. von Stechow & G. Wienold (Eds.), *Lexical knowledge in the organization of language* (pp. 301–340). Amsterdam: Benjamins.