Languages clearly differ in the ways in which they describe spatial location and motion. These cross-linguistic differences have given rise to the question of whether the corresponding nonlinguistic systems that process motion and location are affected by linguistic variation. This chapter discusses a series of experimental findings addressing this question. The independence of nonlinguistic spatial cognition from linguistic encoding preferences is demonstrated. The chapter concludes by considering implications for models of the relationship between language and thought.
1 Language and Thought

According to classical theories of language and cognition, human thought is characterized by strong universal commonalities built around notions such as object, space, agency, number, time, and event (H. Clark, 1973; G. Miller and Johnson-Laird, 1976). Languages select from this prelinguistic conceptual repertoire the concepts that become encoded in their lexical and grammatical stock. Language acquisition, on this view, is a mapping process in which the learner needs to figure out which sounds in the language spoken in the environment correspond to which concepts from the ones already in the mind (Fodor, 1975; Gleitman, 1990; Pinker, 1994).

Within cognitive science this view has recently been challenged by several commentators, who propose that language may not simply reflect, but also shape, underlying cognitive representations (Boroditsky, 2001; Bowerman and Levinson, 2001; Brown and Levinson, 1993a, 1993b; Gentner and Goldin-Meadow, 2003; Imai and Gentner, 1997; Lucy, 1992; Pederson, Danziger, Wilkins, Levinson, Kita, and Senft, 1998; for precursors of this view, see Whorf, 1956; see also chapters by Leslie, Gallistel and Gelman, and Margolis and Laurence in this volume for critical discussion). On this argument, since human beings need to share information about the world through the use of language, humans’ attention during perceptual and cognitive tasks may end up focusing on those dimensions of experience that can most naturally find their way into their native language. If true, this possibility entails that speakers of different languages could end up paying attention to different categories as they experience events in the world and that, in learning language, children may allocate more attention to those concepts that are systematically encoded in their native tongue (see Choi and Bowerman, 1991; McDonough, Choi, and Mandler, 2003; Gentner and Boroditsky, 2001).

These cognitive-relativistic proposals are motivated by the fact that languages differ substantially in their stock of lexical and grammatical semantic distinctions: Number is marked on nominal and verbal expressions in German but not in Chinese; source of information (e.g., whether the speaker witnessed a past event directly or not) is marked morphologically in Turkish but not in Greek; tense is marked (on the verb) in French but not in Thai; and so on. This variation at minimum means that speakers should be able to assemble the appropriate semantically-syntactic information into linguistic strings during language production; in other words, from the earliest stages of speech planning, speakers should be able to construct language-specific conceptualizations of the information they want to convey to listeners (a process that has been called “thinking for speaking” (Slobin, 1996; see Levelt, 1989). But beyond these demands of verbalization, relativistic views take language to affect cognitive processes in much deeper ways. Even though several types of such accounts exist in the literature, they all share the basic assumption that linguistic-semantic categories shape the cognitive processing of events and objects even in situations that involve no verbalization (e.g., Boroditsky, 2001; Lucy, 1992; Pederson et al, 1998; Levinson, 2003).

How can language affect nonlinguistic thought processes? One possibility is that linguistically encoded semantic categories affect the relative accessibility of the corresponding cognitive distinctions (the salience hypothesis). For instance, among the many ways of construing a scene, those that are relevant for linguistic encoding may become “privileged” for memory and categorization. Language, by virtue of being continuously used throughout one’s life, would thus come to affect an individual’s “habitual patterns of thought” (Whorf, 1956) by channeling the individual’s attention toward certain distinctions and away from others. These effects should increase in potency during development: As children gradually acquire their native language, they will become “infected” with its semantic categories. Even though such linguistic...
effects may not be permanent, and can potentially be overridden by other cognitive factors, they can act as a lens on nonlinguistic cognition (Gentner and Goldin-Meadow, 2003).

A subtly different (and stronger) hypothesis maintains that language can have broader reorganization effects on cognitive architecture. On this view, communication pressures require humans to store information about the world in a format that is compatible with language-specific demands so that it can be recovered and quickly converted into an utterance, if needed. According to this position, since languages need to encode different features of objects and events, the mental representation of objects and events will be skewed to encode these features—and ultimately, several cognitive systems that interface with language, including memory, reasoning, and decision-making, will conform to language-specific patterns. Thus “the need to output language coded in specific semantic parameters can force a deep-seated specialization of mind” (Levinson, 2003, p. 291). On this cognitive streamlining hypothesis, the alignment of cognitive resources with linguistic encoding preferences may create striking (and permanent) cognitive discontinuities between speakers of different languages (for further discussion, see Levinson, 2003; Pederson et al., 1998; Majid, Bowerman, Kita, Haun, and Levinson, 2004; and section 4 below).

Given the importance of the relationship between language and thought, and the controversy of the issues surrounding it, it is worth considering these two hypotheses more closely, examining what they can and cannot mean, and evaluating the experimental evidence pertaining to them. The purpose of this chapter is to do just that, focusing especially on the domain of motion and space. Together, the studies we will review compare performance of linguistic and nonlinguistic spatial tasks by speakers of different languages and across various age groups to test whether certain spatial concepts are privileged over others, which ones are natural/universal, and whether the way spatial concepts get encoded in language affects nonlinguistic spatial cognition. We cannot review here the large literature examining potential effects (and non-effects) of language on thought in other domains (but see the contributions in Bowerman and Levinson, 2001; Gentner and Goldin-Meadow, 2003; Gumperz and Levinson, 1996; and for a review, Gleitman and Papafragou, 2005).

1.1 Motion and Location in Language and Thought

Motion and space form an ideal empirical area for studying the links between cognition and language. First, the representation of motion and space is a fundamental human cognitive ability (Pick and Acredolo, 1983; Stiles-Davis, Kritchevsky, and Bellugi, 1988; Emmorey and Reilly, 1995; Hayward and Tarr, 1995; Newcombe and Huttenlocher, 2003; Carlson and van der Zee, 2005; and many others). Some of this spatial-mechanical conceptual machinery is already available early in life: Prelinguistic infants can parse dynamic events into objects and actions (Baldwin, 1991; K. Wynn, 1996); discriminate spatial relations within events (Casasola, Cohen, and Chiarello, 2003); and categorize abstract relations of spatial constructs (Casasola and Cohen, 2002; Quinn, 1994).

Second, motion and location are important cross-linguistic typological domains. So far as we know, all languages have “where” questions, usually with a single morpheme that conflates motion and static location (Ulltian, 1978). Furthermore, expressions of how objects move or are located in space cross-linguistically draw upon a recurrent set of distinctions (including topological notions such as containment, support, attachment, and contiguity; geometric notions such as axial structure of the moving/located object; and so on; Talmy, 1985; Landau and Jackendoff, 1993). Despite these broad similarities, the linguistic encoding of motion and space is characterized by intense typological variability, as we will discuss in detail below.

Third, motion and space terms are acquired early by language-learning children around the world (Berman and Slobin, 1994). For instance, very young children generalize spatial words such as up, down or back very
fast, on the basis of very few exposures (Smiley and Huttenlocher, 1995; McCune-Nicolich, 1981).
Furthermore, children conform to language-specific patterns as they start acquiring the spatial terminology of their language (Choi and Bowerman, 1991).

In what follows, we discuss two specific case studies, motion (section 2) and location (or frames of reference, section 3), to address what is universal and what is language-dependent about spatial cognition. Each case study focuses on one of the two specific hypotheses we have laid out previously about the relationship between linguistic and conceptual spatial representations.

## 2 Motion Events

It is generally recognized that the ability to talk about motion is supported by a set of “natural,” probably universal, event primitives which guide the cognitive partitioning of motion events: These include PATH, or trajectory (e.g., entering or exiting); MANNER (spinning or rolling), and CAUSE (someone bounces a ball versus a ball bounces; Talmy, 1985; Landau and Jackendo, 1993). Some of these concepts appear early on: We know that infants in the first year of life detect changes in the path and manner of events (Pulverman, Sootsman, Golinko, and Hirsh-Pasek, 2003), and find the invariant path and manner in actions (Pruden, Hirsh-Pasek, Maguire, Meyers, and Golinkoff, 2004). We also know that these motion primitives structure the gestural system of deaf signers (Zheng and Goldin-Meadow, 2002; Senghas, Kita, and Özyürek, 2004).

Despite these commonalities, both the way motion primitives are lexicalized in spatial vocabularies and the way these primitives are conflated into sentential structure vary considerably cross-linguistically (Talmy, 1975). For instance, in languages such as English, German, Russian, and Chinese, manner information is usually encoded in the verb (e.g., “The bottle floated ...”) and path information appears in particles, prepositional phrases, or other nonverbal modifiers (“... into the cave”). By contrast, in languages such as Greek, French, Spanish, and Turkish, verbs often encode path information (“The bottle entered the cave ...”), while manner may be encoded in modifiers or omitted altogether (... (floating)). Furthermore, the distribution of manner verbs in the last group of languages is quite constrained: In Greek, for instance, most manner verbs cannot combine with a modifier which denotes a bounded, completed path (as in “A boat sailed to the island”)—a path verb needs to be used instead (“A boat arrived at the island”). This boundedness constraint leads to higher use of path verbs in Greek compared to English (a similar constraint is found in several languages of the Greek group; see Aske, 1989; Jackendoft, 1990; Slobin and Hoiting, 1994; Levin and Rapoport, 1988).

These cross-linguistic differences in the verbal encoding of motion have been confirmed in psycholinguistic studies with both adults and children (Allen, Özyürek, Kita, Brown, Turanli, and Ishizuka, 2003; Choi and Bowerman, 1991; Naigles, Eisenberg, Kako, Highter, and McGraw, 1998; Özçalişkan and Slobin, 1999; Sebastián and Slobin, 1994; Slobin, 1996, 2004). Could such “packaging” differences impact the nonlinguistic representation of motion events? If manner of motion is a prominent lexicalization feature of English and similar languages, the salience hypothesis would predict that manner would be more likely to determine how motion events are categorized or remembered in those languages than in languages where it is less prominent (e.g., Greek).² From a developmental perspective, the salience view entails that the dimension of manner may enjoy higher accessibility in the minds of English learners even in situations that do not explicitly involve linguistic communication (Bowerman and Choi, 2004).
2.1 Testing the Salience Hypothesis

In a series of studies, Papafragou, Massey, and Gleitman (2002, 2006) set out to test whether motion event cognition was affected by the way motion was encoded in English versus Greek. On the basis of elicited descriptions of motion scenes, the studies first confirmed the manner/path asymmetry in the speech of Greek- versus English-speaking children and, much more strongly, Greek- versus English-speaking adults. Papafragou et al. (2002), then went on to compare their English- and Greek-speaking subjects on memory of path or manner details of motion scenes. Children and adults were presented with a set of black-and-white drawings depicting motion events (e.g., a frog jumping off a turtle). In a second session a few days later, subjects were given another set of drawings and were asked whether they were the same as or different from the ones they had seen before. The drawings in the second set were the same as the originals, path variants (a frog jumping onto a turtle) or manner variants (a frog falling off a turtle; see figure 14.1). Despite the asymmetry in verbally encoding the events, memory results showed that English and Greek speakers were equally good at detecting manner or path changes to the original pictures: Strikingly, memory accuracy in this task was closely matched across the different age groups in the two linguistic populations.

Figure 14.1  Stimuli for memory experiment (Papafragou et al., 2002). Panel (a) represents a target stimulus; panel (b), a path change; and panel (c), a manner change.

A second experiment tested sensitivity to manner/path distinctions as a basis for categorizing motion events. Adults and eight-year-old children were shown sets of motion events consisting of a sample event (e.g., a man stumbling into a room) and two variants: a manner variant (a man walking into a room) and a path variant (a man stumbling down the stairs). Subjects were asked to choose the variant where the agent was “doing the same thing” as in the sample. Again, English and Greek speakers did not differ from each other in terms of classification for path and manner—and this was true for both the eight-year-olds and the adults.

The same set of studies suggests that the mental representation of motion is independent of linguistic naming even within a single language. Papafragou et al. (2002) divided their English- and Greek-speaking subjects' verbal descriptions of motion according to whether they included a path or manner verb, regardless of native language. It was found that verb choice did not predict choice of path/manner as a basis for remembering or categorizing motion scenes. Naming and cognition, then, are distinct under these conditions: Even for speakers of a single language, the linguistic resources mobilized for labeling do not faithfully mirror the cognitive resources mobilized for cognitive processing (e.g., memorizing, classifying, reasoning, etc.).

Similar results have been obtained for Spanish versus English by Gennari, Sloman, Malt and Fitch (2002). In their study, English and Spanish speakers' descriptions of motion clips were compared to subjects' performance in two nonlinguistic tasks: recognition memory and similarity judgments. The study varied
whether subjects described the events verbally during initial event apprehension or not. No effect of language was found in the recognition memory task after either linguistic or nonlinguistic encoding, and in the similarity task after nonlinguistic encoding. However, when subjects were asked to perform a similarity judgment task immediately after having described motion events in their native language, it was found that the linguistic labeling affected their choices. These results support the conclusion that linguistic and nonlinguistic motion representations are dissociable. They also show that when language-specific regularities are made salient in an experimental context, they can mediate subjects’ performance in certain tasks: Categorization seems to be especially vulnerable to such linguistic intrusions (see also Finkbeiner, Nicol, Greth, and Nakamura, 2002).

One question which arises from these studies is whether the manner of motion is monitored during communication even in languages that do not systematically encode it. Interestingly, subsequent analysis of the linguistic data in Papafragou et al. (2002) revealed that Greek speakers were more likely to include manner of motion in their verbal descriptions when manner was unexpected or noninferable, while English speakers included manner information regardless of inferability (Papafragou et al., 2006). For instance, Greek speakers were more likely to include a manner modifier or other manner element to describe a scene where a man was running versus walking up the stairs (presumably because walking is the typical, expected way for the man to move, while running is atypical/noninferable). No such difference existed in English, since manner is almost always encoded in the verb. The Greek data suggest that manner of motion information, even if not prominently encoded in participants’ responses, is nevertheless not lost: Speakers may monitor such harder-to-encode event components and choose to include them in their utterances when they are especially informative. This finding offers further support for the conclusion that language productions do not faithfully represent cognitive representations but are only a pointer to richer underlying conceptual structures.

The set of studies reported here converges on the conclusion that the perception and conceptual organization of motion events are independent of language-specific encoding preferences. This conclusion is supported by other studies that have compared linguistic and cognitive spatial representation. For instance, Munnich, Landau, and Dosher (2001) studied the on/above spatial contrast which is encoded by different prepositions in English but is not grammaticalized in Korean: They found that despite this difference, English and Korean speakers had equally good memory for the relevant spatial locations. Similar effects have been observed in object categorization: Malt, Sloman, Gennari, Shi, and Wang (1999) found that despite differences in the vocabulary used by English, Spanish, and Chinese subjects for everyday containers such as bottles, jars, jugs, and boxes, similarity judgments for different containers by speakers of the three languages did not differ much (if at all). Again naming and nonlinguistic categorization seem to diverge, with perceptual resemblances, historical factors, and arbitrariness all contributing to naming decisions in a specific linguistic community.

2.2 Linguistic Motion Categories as On-line “Attention Magnets”?

The studies reviewed so far have addressed the possibility that cross-linguistic differences in motion event encoding affect the salience of event components in memory and categorization. But one might object that in such temporally unconstrained tasks, language effects may be masked or overridden. A different way of approaching the salience hypothesis is to ask whether one’s native language might shape the on-line allocation of attention to aspects of scenes. If true, this would entail that event perception, or at least the earliest postperceptual moments of event apprehension, can be affected by language.

This hypothesis is distinct from the idea, common in models of language production, that language-specific demands on the formulation of messages have become automatized in adult speakers, and shape the preparation of encodable messages even before the activation of specific lexical items (Griffin and Bock,
2000; Levelt, 1989; see the discussion of Slobin’s notion of “thinking for speaking” in section 1). For instance, it is quite possible, in fact likely, that manner and path event components are in differential states of conceptual readiness in the minds of Greek and English speakers immediately prior to their uttering a description of a motion scene. What is more controversial is whether speakers of English and Greek might attend differentially to such event components even when they are not engaged in verbal communication.

One way of testing this proposal would be to monitor people’s eye movements as they inspect motion scenes during either a linguistic or a nonlinguistic task. We know that where people look, reflects pretty accurately what they are getting ready to encode linguistically (Griffin and Bock, 2000), or more generally what people are focusing their attention on under normal viewing conditions. If language affects attention allocation early in event perception, eye movements of speakers of different languages should differ when they inspect events, regardless of whether they are engaged in communication or not. To address this issue, Papafragou, Hulbert, and Trueswell (2006) recorded eye movements of native Greek and English speakers as they watched short (three-second) clip art animations depicting motion events. Stimuli included both complex, bounded events (e.g., a man skating to a snowman; see figure 14.2), and simple, unbounded ones (e.g., a man skating). At the end of each clip, a beep was heard and the last frame of the clip froze on the screen for two seconds. Participants were asked to perform either a linguistic task (i.e., describe aloud the video after the beep) or a nonlinguistic task (i.e., study the video for a later image recognition test).

Recall that Greek and English differ in that only English allows the expression of manner in the verb in bounded events (“The man is skating to the snowman”). In Greek, the same event typically needs to be described by a path verb (e.g., “The man is approaching the snowman, skating”). Both languages can use manner verbs to describe unbounded events (“A man is skating”). If cross-linguistic differences in motion encoding affect motion perception, we would expect English speakers to be more likely than Greek speakers to focus on the manner of motion early and consistently in both the linguistic and the nonlinguistic task. But if event perception is independent of language, we should see differences between English and Greek speakers only in the linguistic task (and only in bounded events, where the languages clearly differ).

As anticipated, participants’ verbal responses showed that the English group was much more likely to produce manner verbs than the Greek group in bounded events, but that the difference was much smaller in
unbounded events. Unsurprisingly, analysis of eye movements revealed that inspection of the unbounded events did not differ between English and Greek speakers in either the linguistic or the nonlinguistic task.

Turning to the bounded events, the eye movement data indicated some interesting asymmetries between Greek and English speakers in the linguistic condition. Specifically, as soon as motion began, participants looked to the regions corresponding to event components that their language typically encodes in verbs: Greeks were more likely to look first to the end point of the motion path (e.g., the snowman) and English speakers to the manner (e.g., the skates). After the beep, when people prepared to describe the bounded events, these eye movement patterns were repeated. Language-specific verb lexicalization preferences, therefore, affect how events are processed for purposes of communication—a particularly clear case of “thinking for speaking.” Importantly, the overall time spent looking at manner or path did not differ between the two languages; what differed was when each of these two event components was attended to (in accordance with verb typologies).

Remarkably, in the nonlinguistic condition, eye movements during inspection of bounded events were nearly identical for both language groups. In other words, when people inspected events freely, the way they allocated attention was similar regardless of their native language. This is a clear demonstration of the independence of event perception from language-specific encoding preferences.

The nonlinguistic condition yielded one further important finding for bounded events that adds complexity and subtlety to the picture just sketched. At the end of each trial (after the beep), when the clip had finished and its last frame remained on the screen, people focused their gaze on those motion components not typically encoded in the main verb in their language (Greek speakers were concerned about manner and English speakers about path). A reasonable explanation of why this occurred is that participants were preparing for the upcoming memory task and were directing their attention to those aspects of the scene that did not fit the encoding preferences in their language. This indicates that participants were somehow converting the scene into a linguistic representation to support its later recall. We know from much prior research that such linguistic recoding does occur in many memory tasks (Conrad, 1964; Baddeley, 2003). What is novel here is that participants seem to be focusing on aspects of events that do not lend themselves to easy verb encoding. This finding suggests a role for language in storing motion events in memory that needs to be explored further. This role, however, is very different from what is straightforwardly predicted by the salience hypothesis, since linguistically privileged event components do not seem to attract special attention in preparation for memory storage and retrieval.

The eye-tracking methodology has great potential for illuminating the interplay of linguistic and cognitive factors in further tasks (such as speeded memory judgments) and with different populations (children as well as adults). So far, this technique has led to data offering a clear distinction between processes underlying inspection of a scene for purposes of linguistic description versus memorization. These data confirm the idea that language-specific demands on message formulation affect how people parse scenes prior to encoding them linguistically. Furthermore, they show for the first time that cross-linguistic differences can impact speech planning in terms of when people extract and attend to information from an ongoing scene. However, our data show that these planning effects do not percolate to nonlinguistic tasks such as event apprehension. People seem to view dynamic events in similar ways regardless of the language they speak.
We now turn to a second empirical area, the linguistic description of location and orientation. As several commentators have pointed out, languages differ widely in how they locate objects in space (Brown and Levinson, 1992, 1993a, 1993b; Levinson, 1996, 2003; Levinson et al., 2002; Majid et al., 2004; Pederson et al., 1998). For instance, English typically uses an egocentric, bodily defined coordinate system to locate objects or give directions (e.g., “The ball is to my left” or “Turn right”). Many, perhaps a third, of all languages, however, lack expressions for familiar spatial notions such as “left” and “right.” One such language is Tseltal Mayan, spoken in Tenejapa and other communities in the highland region of Chiapas in Mexico. Although Tseltal has terms for “left” (xin) and “right” (wa’el), they are used extremely infrequently, and only to reference body parts, not regions outside the body (e.g., to one’s left or right). In order to talk about spatial arrangements or directions, Tseltal speakers use a system of geocentric coordinates based on the overall inclination of the land (alan “downhill” and ajk’ol “uphill”). These terms have been extended to refer to the north-south axis, and can thus be used even when one is on flat terrain (e.g., “The ball is uphill/south from me”). For locations on the horizontal, orthogonal to the uphill-downhill axis, Tseltal uses ta jejch (“crosshill”).

These impressive cross-linguistic differences in encoding location and orientation have been shown to correlate quite closely with performance on various nonlinguistic tasks involving spatial reasoning (see references above). In one of these tasks, Tseltal-speaking Tenejapans were compared to speakers of Dutch, a language which—like English—has “left”/“right” vocabulary. On a typical trial, participants were presented with a card on which were a small circle and a large circle of different colors (see figure 14.3). The card was removed, and after a short delay participants were rotated 180° to face an array of four cards identical to the first one but arranged in distinct orientations. The task was to choose the card that was identical to the original. As is obvious from the figure, this task had both an egocentric solution (participants could choose the card with the large dot to their right) and a geocentric solution (participants could pick the card with the large dot to the south). Dutch subjects give consistently egocentric responses in this task, while Tseltal subjects, to the extent that they gave consistent responses, preferred geocentric solutions over 80 percent of the time.

Other tasks implementing the same logic gave similar results. For instance, in an experiment testing recognition memory, participants observed a toy man “walking” along and making rectangular turns. After rotation, they were asked to find the path traversed by the toy man on a map or maze that contained numerous possible paths, including the one observed, regardless of whether it was coded egocentrically or geocentrically. Again, Tenejapans preserved the absolute direction of the motion path, while the Dutch...
responded egocentrically, on the basis of the relation of the arcs of the path to themselves. Taken together, these data have led the above group of researchers to the conclusion that Tseltal speakers do not use left-right distinctions in their habitual reasoning about space, but instead rely on an absolute (viewpoint-independent) coordinate system (see, e.g., Pederson et al., 1998). Tseltal-speaking Tenejapans have been credited with a “learned ability to maintain fixed bearings at all times” (Levinson, 2003, p. 168), a sort of mental compass which enables them to calculate external coordinates on an array or path regardless of their own position. In further research, asymmetries similar to those in the Dutch-Tenejapan case have been uncovered for several other language groups (Pederson et al., 1998).

Crucially, these data have been interpreted as evidence for the conclusion that linguistic frames of reference are an organizing force for the way space is encoded in other cognitive faculties. These frames, the argument goes, are not logically equivalent, since they capture different kinds of information (if I know that the cup is to your north, I cannot tell without additional information whether it is also to your left or your right, or in some other position). The fact that language communities make choices about preferred frames of reference forces speakers to encode spatial configurations in the linguistically preferred mode every time they process spatial information, in case they need to talk about such configurations later in time:

Such [linguistic] restrictions place a bottleneck on the entire system of [spatial] representations—if we are to talk about what we see and feel and remember, we must make sure that those representations are consistent with the available linguistic ones, or can be converted into them.... Thus the facts that (a) frameworks are not freely convertible, (b) languages may offer restricted frameworks as output, and (c) it may be desirable to describe any spatial experience whatsoever at some later point, all conspire to require that speakers code spatial experiences at the time of experience in whatever output frameworks their dominant language offers. (Levinson, 2003, pp. 60–61)

Once a language has opted for one of these frames of reference and not the other, all the systems that support language, from memory, to reasoning to gesture, have to provide information in the same frame of reference. (ibid., p. 290)

More than simply promoting the salience of certain categories over others, the linguistic choice of frames of reference is thus assumed to streamline cognitive resources so that they respond most efficiently to the pressures of rapid communication.
3.1 Testing the “Cognitive Streamlining” Hypothesis

The experimental findings summarized in the last section demonstrate a clear congruence between linguistic spatial encoding and nonlinguistic spatial reasoning in different communities. But whether language is responsible for the molding of nonlinguistic spatial systems is less clear from the data: Alternatively, it could be that the very same reasons that lead members of a specific community to choose a preferred way of reasoning about space (e.g., in the case of Tenejapans, rural environment, stability and availability of local landmarks, etc.) determine a preference for a linguistic coordinate system (here, a geocentric one). A different possibility is that language mediates the solution of the rotation tasks, thereby turning them into linguistic problems. Notice that these tasks are ambiguous (the instruction to make two arrays “the same” allows for both geocentric and egocentric responses). It is possible that in deciding what counts as “the same” spatial array or path, participants in such tasks fell back on linguistic terminology and considered what would be the most likely interpretation of these instructions, given the linguistic and cultural norms of the community (see Li and Gleitman, 2002, for discussion). Subsequent research has shown that the spatial reasoning preferences for English speakers are not consistently egocentric, but depend on various environmental factors (Li and Gleitman, 2002). A question that arises is whether, and to what extent, Tseltal speakers’ spatial reasoning is also flexible, despite the lack of egocentric terminology. Specifically, it is of interest to determine whether Tseltal speakers can use egocentric reasoning when presented with clear, unambiguous spatial tasks.

We decided to probe further into the scope and potential limitations of Tseltal spatial reasoning in a series of experiments that was inspired by prior studies but had a unique (either egocentric or geocentric) solution (Li, Abarbanell, and Papafragou, 2005). Our tasks kept the use of the rotation technique (which usefully dissociates egocentric and geocentric encoding of spatial arrays) but added disambiguating (nonlinguistic) cues. The logic behind these manipulations was simple: If linguistic frames of reference place constraints on the accessibility or even the availability of spatial reasoning resources, egocentric solutions to these tasks should be more difficult for Tseltal speakers, compared to the geocentric solutions favored by their linguistic system.

Participants in our studies were (mostly monolingual) adult speakers of Tseltal recruited during fieldwork in Tenejapa. The first of our tasks was a modification of the cards task we have described above (see figure 14.3). Participants inspected a card with two circles (e.g., a yellow and a green circle) that was placed on a table in the experimental scene. They then placed the card in a box, closed the lid, and, turning 180°, transferred the box onto an identically oriented table at the other end of the room. The tables were placed such that the participants’ left and right were aligned with the uphill-downhill (north-south) axis. The participants’ task was to identify, without opening the box, which one of four identical but distinctly oriented cards on the second table was “the same” as the card in the box.

Participants were randomly assigned to either an egocentric (EC) or a geocentric (GC) condition. In the first case, they rotated together with the box and went over to the second table; in the second case, the participant rotated, but the box itself did not (i.e., the participant rotated before he/she picked up the box at the first table). This simple difference was expected to prompt different responses on the part of the participants. This expectation was borne out there were 85 percent versus 74 percent of correct responses in the EC and the GC condition, respectively. Furthermore, there was no difference between the two conditions: Participants were equally successful with the “linguistically dispreferred” and the linguistically dominant frames of reference.

A second task adapted the maze task discussed earlier. Participants had to memorize the path traversed by a ball in a maze (see figure 14.4). Then the maze was covered and carried to a second table, where the participant had to re-create the ball’s path. As with the cards task, in the EC condition, participants held the covered maze as they rotated, so that the maze rotated with them; in the GC condition, participants rotated
themselves to face the second table before picking up the covered maze, so that the maze itself was not rotated. We anticipated that this step would tacitly encourage participants to encode in memory the motion path previously demonstrated in the maze in EC versus GC terms, so that they could accurately re-create it on the second table.

**Figure 14.4** Sample trial for the maze task. (Li et al., 2005)

Motion paths in this experiment varied in terms of complexity, with one-, two-, and three-leg paths presented to participants in this order (see figure 14.4 for an example of a two-leg path). Previous commentators had hypothesized that higher memory load in spatial reasoning tasks leads participants to resort to their “unreflective, natural mode of coding” (Levinson, 2003, p. 199). If true, this would yield higher error rates for the EC condition compared to the GC condition as the number of legs increased. What we found instead was exactly the opposite: Tseltal speakers were much better at the EC than the GC condition in the maze task (92 percent versus 62 percent of paths were traced correctly in their entirety, respectively). Even more strikingly, as path complexity increased across trials, the Tenejapans’ success in the EC condition remained relatively stable (100 percent, 96 percent, and 80 percent in one-, two-, and three-leg paths), while in the GC condition it declined dramatically (92 percent, 60 percent, and 35 percent, respectively). Once again, these data show convincingly that Tseltal speakers can and do use egocentric coordinates to solve spatial tasks; importantly, they also show that such coordinates are, in fact, more accurately computed, especially under increased memory demands, than geocentric coordinates—exactly the opposite of what one would expect on the basis of the linguistic frames of reference systems found in Tseltal.

These conclusions were bolstered by results of a third task which tested participants’ search abilities. Subjects sat on a swivel chair while a coin was hidden by the experimenter in one of two boxes on either side (left-right/north-south) of the chair. The participant was then blindfolded, spun around 360° plus an additional 90°, 180°, 270°, or 360°, and then allowed to look for the coin after the blindfold had been removed. In the EC condition, the two boxes were fastened to two spokes attached to the chair so that they rotated with the participant; in the GC condition, the boxes were placed on the floor and did not change place. Unlike previous experiments, each participant was tested in both the EC and the GC condition. We found that participants succeeded in retrieving the coin on the first try 92 percent of the time in the EC condition but only 80 percent in the GC condition (a statistically significant difference). This happened despite the fact that cues for the geocentric encoding of the scene (e.g., details about the testing room, local landmarks) were readily available and could easily have been used.

Overall, these studies show that Tseltal speakers, when given implicit cues that body-centered (left-right) distinctions are needed to solve a spatial task, use these distinctions without problems. On certain tasks,
performance with such body–centered distinctions is better than performance with absolute systems of orientation that correspond more closely to the preferred linguistic systems of encoding space in Tseltal. Tseltal speakers’ switch to egocentric reasoning seems to happen flexibly and without much training, despite what one might expect on the basis of language-specific frame-of-reference choices and prior empirical results. We take this as another demonstration of the independence of spatial reasoning from linguistic encoding preferences. The linguistic and nonlinguistic representations of space, even though correlated, are distinct and dissociable. If anything, the linguistic representation of space underrepresents the cognitive representation systems that underlie spatial thought (see also Gleitman, Li, Abarbanell, Gallistel, and Papafragou, in preparation, for further discussion and experimentation).

3.2 Semantics and Cognition: Underdeterminacy Versus Isomorphism

The conclusion of the last section raises an issue which lies at the core of many recent (and several old) debates about language and thought, so we will consider its implications a little further before we conclude this chapter. Traditionally, in the cognitive sciences, language has been considered a privileged entry point into the nature of conceptual representations (Jackendo, 1990). More strongly, it has sometimes been assumed that linguistic-semantic and conceptual representations are simply identical (see, e.g., Jackendo, 1996; Langacker, 1987). The “cognitive streamlining” hypothesis has to agree with this general position: In order for language to have reorganizing effects on other cognitive systems, linguistic-semantic representations need to be equivalent, or at least isomorphic, to conceptual representations.

However, there are several reasons for believing that linguistic and cognitive representations cannot be equivalent (see Fodor, 1975; Pinker, 1994; Sperber and Wilson, 1986; Gleitman and Papafragou, 2005, for discussion). First, there are several phenomena specific to language that have no counterparts in thought. Words can be ambiguous (the English word “seal” corresponds to several distinct concepts), whereas concepts are not. Similarly, the meanings of several linguistic expressions are impoverished and broad, and need support from context to be fleshed out into a complete thought (in the sentence “He saw her sitting there,” he, her, and there all require extralinguistic information to be able to pick out specific people and places). In fact, almost any utterance relies on context to convey the exact meaning the speaker has in mind; as a result, even though the meanings of the words do not change, their interpretations vary from utterance to utterance (“The room is hot” and “The oven is hot” convey different construals of the adjective “hot”). Finally, everyday communication gives rise to conversational inferences which ensure that what a sentence semantically (or literally) means and what the sentence conveys when used by a specific speaker in a specific situation will never be quite the same (a famous example is Yogi Berra’s “Nobody goes there anymore—it’s too crowded”). This and related evidence supports the conclusion that linguistically encoded meaning is only a pointer to the thought the speaker had in mind and wanted to communicate (or, as it is sometimes put, language underdetermines thought content).

Far from being incidental or peripheral, the underdeterminacy property is in fact a design feature of human language (for detailed discussion, see Carston, 2002). For reasons of economy, both speakers and listeners make heavy use of inferential devices which make it possible for relatively complex thoughts to be conveyed rapidly and efficiently by simple and short utterances (e.g., “Tomorrow is another day”). To succeed in this game, speakers have to constantly select how much of their thinking to encode in words and how much to leave up to listeners to infer, and hearers have to take into account the speakers’ knowledge and intentions in inferentially expanding upon what was said. It is this principled, joint mind-guessing effort that allows conversation ever to get off the ground (Grice, 1975; Sperber and Wilson, 1986; Clark, 1992; Bloom, 2000).

Given the powers of the human inferential machinery to go beyond what is linguistically encoded in a sentence, it appears unlikely that linguistic-semantic resources would shape mental architecture in the strong way the cognitive streamlining position maintains (Papafragou, 2004). Paradoxically, proponents of
this view sometimes acknowledge that language and cognition cannot rely on the same kind of representation (for the reasons just discussed) but insist that, in the end, the corresponding representations converge (Levinson, 2003).

Even setting aside the underdeterminacy argument, the claim that (language-specific) lexical-semantic categories become the categories of thought faces several difficulties. Clearly, not all lexical differences among languages cause cognitive reorganization, so one would have to specify which ones do. Most important, even when cross-linguistic differences seem pervasive, lexico-semantic structure does not take over mental representations. Comparative experimental evidence from domains as disparate as color (Heider and Oliver, 1972), counterfactual reasoning (Au, 1983), and geometry (Dehaene, Izard, Pica, and Spelke, 2006) shows that linguistic distinctions (or lack thereof) do not skew (nonlinguistic) cognitive distinctions. Together with our own data on Tseltal, these findings support the conclusion that linguistic representations do not streamline the cognitive representations individuals possess and bring to bear on their interactions with the world.

4 Concluding Remarks

The cross-linguistic investigation of spatial encoding and its interface with nonlinguistic spatial representations is only beginning, and there is still much controversy surrounding the precise form and properties of the interface. Nevertheless, it appears that, at least for the cases reviewed here, the picture of the interface that emerges is rather traditional. Linguistic semantic-syntactic structures underrepresent the cognitive representation of space and motion scenes. Attempts to test predictions of relativistic accounts have turned up mostly negative findings. Despite differences in how space and motion scenes are encoded cross-linguistically, a series of experiments revealed remarkable similarities in the way movement and location are perceived and processed.

Nevertheless, some aspects of our findings suggest a role for language, even though a more limited one than the one envisaged in recent relativistic accounts. Language can provide support for memory by allowing the recoding of spatial information into linguistic terms. This, in turn, can make speakers sensitive to kinds of information that can (and cannot) be encoded linguistically in a specific language, and impact on-line attention allocation accordingly. Linguistic intrusions into memory tasks have been uncovered in our eyetracking motion study discussed earlier, and need to be explored further. The eye-tracking technology seems a particularly apt tool for studying such effects, since it allows direct insight into the process of how attention is distributed onto elements of a scene during both linguistic and nonlinguistic tasks.

Even though this chapter has focused on a single empirical area, we take these results to be instructive for the investigation of the language-thought interface in other domains. The findings on spatial cognition themselves, of course, do not preclude the presence of linguistic effects on other aspects of thought; however, our discussion may help in the understanding of the nature and scope of potential effects, and set some boundaries on what sorts of effects can reasonably be expected, and what mechanisms operate at the interface of language and cognition.

Two important questions remain open and need to be addressed in the next stages of research into the relationship between language and thought. The first concerns what counts as a cross-linguistic difference. Given that vocabularies across languages can differ in innumerable ways, most researchers have focused on grammatical (rather than lexical) candidates for language-on-thought effects. Even there, emphasis has typically been placed on differences in what is an obligatory (rather than optional) semantico-syntactic category, or sometimes a habitual/frequent (rather than an infrequently used) distinction. Each of these levels leads to different generalizations about the kinds of linguistic effects expected, and there is no
consensus about whether, say, any two obligatory grammatical distinctions (e.g., tense and number in English) have equal potential for causing linguistic effects on thought.

A related, but deeper and more difficult, question is what counts as a linguistic-semantic universal. In this chapter, we have taken the perspective (common to most, but by no means all, commentators in the field) that there are universal notions of spatial structure and that these notions impact the intertwined processes of learning and processing language. So far, proposals for spatial primitives have been based on a close examination of the English prepositional system (Landau and Jackendoff, 1993), with a certain amount of analysis of other languages (e.g., Talmy, 1975, 1985). However, more detailed typological work on a wider sample of languages remains to be done in order to confirm and modify these proposals. Further valuable evidence about the inventory of spatial universals can come from cross-linguistic comparative acquisition work with infants and young children which is only beginning (see Choi, McDonough, Bowerman, and Mandler, 1999; McDonough et al., 2003; Pulverman, Golinkoff, Hirsh-Pasek, and Jackson-Maldonado, 2005). The challenge for the next stages of research is to pursue this integrated approach so as to identify more precisely what is universal and what is language-specific about the way perceivers view and talk about events in the world.

Notes

1. We will not discuss here strong versions of the language-on-thought hypothesis, according to which language is a precondition of thought (i.e. it creates underlying cognitive structures). This is a view associated with some of Whorf’s (1956) writings, but has long been discredited within modern cognitive science (Pinker, 1994).

2. More generally, on this view, “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 and Boroditsky, 2001, p. 247).

3. We know that language-specific encoding preferences can affect conjectures about the meaning of newly encountered lexical items. Adult speakers of Spanish and English, when asked to guess the meaning of novel motion verbs, show sensitivity to the lexicalization statistics for motion in the two languages (Spanish speakers prefer path, and English speakers manner, conjectures), as well as the semantic implications of the syntactic frames for motion verbs (Naigles and Terrazas, 1998). Children also seem to be aware of such lexical and syntactic cues when hypothesizing the meaning of novel motion verbs cross-linguistically (Hohenstein and Naigles, 2004; Papafragou and Selimis, 2007). In that sense, typological preferences function as a “zoom lens” for event construal in the context of acquiring linguistic form-meaning mappings (see Brown, 1957; Gleitman, 1990; Naigles, 1990; Fisher, 1996). This, of course, is an effect of language structure on language acquisition, and not a bona fide language-on-thought effect.

4. Using a linguistic elicitation task modeled after Pederson et al. (1998), we confirmed that our target Tenejapan population used predominantly geocentric terms (and failed to produce any left/right terms) in discussing spatial arrangements.

5. It has to be noted that the focus on lexical semantics confounds potential linguistic effects with effects of culture. After all, different communities invent words for objects, actions, and properties they find useful for their purposes (such as terms for sculpture in ancient Greece and terms for hunting in Elizabethan England). It is possible that such terminology could itself create salience effects; these effects, however, are weaker and more transient than the effects proposed by the cognitive streamlining view.