Journal of Psycholinguistic Research, Vol. 31, No. 5, September 2002 (© 2002)

The Role of Language in Memory for Actions

**Matthew Finkbeiner,[1,2] Janet Nicol,[1] Delia Greth,[1]
and Kumiko Nakamura[1,3]**

Languages differ with respect to how aspects of motion events tend to be lexicalized. English typically conflates MOTION *with* MANNER*, but Japanese and Spanish typically do not. We report a set of experiments that assessed the effect of this cross-linguistic difference on participants' decisions in a similarity-judgment task about scenes containing novel animations as stimuli. In Experiment 1, which required participants to encode the stimuli briefly into memory, we observed a language effect; in Experiment 2, which required participants to analyze the same stimuli, but not remember them, the language effect disappeared. Hence, these experiments reveal a task-dependent effect, which, we argue, points to working memory as the source of the language effect observed in Experiment 1 and, potentially, other experiments that have shown a linguistic relativity effect.*

KEY WORDS: language; linguistic relativity; motion events; working memory.

INTRODUCTION

Motion events typically involve some kind of movement from one point to another. Languages differ in their tendency to encode the manner of movement as part of the verb or to express manner in a separate lexical item. Some languages, such as English, contain many verbs that encode the manner of the action (verbs such as *trot, scamper, slither*), whereas other lan-

---

The authors gratefully acknowledge support for this research from a number of sources, including the Faculty Small Grants Program (Office of the Vice President for Research and Graduate Studies, in conjunction with the University of Arizona Foundation, grant NIDCD DC-01409 (as part of the National Center for Neurogenic Communication Disorders, University of Arizona), and the Cognitive Science Program, University of Arizona.

[1] University of Arizona, Tucson, Arizona.

[2] To whom correspondence should be addressed at Communication 302, P. O. Box 210025, University of Arizona, Tucson, Arizona 85721-0025. Fax: 520-626-4300; email: msf@u.arizona.edu

[3] The order of the first two authors was determined by flipping a coin.

0090-6905/02/0900-0447/0 © 2002 Plenum Publishing Corporation

guages, such as Spanish and Japanese, contain fewer such verbs; manner of motion may appear separately from the lexical item expressing motion (Choi & Bowerman, 1991; Kageyama, 1996; Naigles & Terrazas, 1998; Talmy, 2000). For example, the way to express "The boy ran across the street" in Spanish and Japanese is as follows:

(1) El niño cruzo' la calle corriendo.
The boy crossed the street running.
(2) sono otokonoko-ga to Hri-o hashi-Qte wataQ-ta
the boy-NOM street-ACC run-GERUND cross-PAST
The boy crossed the street running.

Note that in both examples the manner of crossing the street is expressed in a separate lexical item from the verb expressing motion (note further that in these particular cases, direction of movement, or *path*, is conflated with motion). Although Spanish and Japanese are distinct from English in terms of lexicalized properties of verbs (Talmy, 2000), the distinction is not absolute: both Spanish speakers and Japanese speakers frequently use verbs that conflate manner with motion. In a recent study, Naigles, Eisenberg, Kako, Highter, and McGraw (1998) observed that whereas English speakers used primarily manner verbs in their descriptions of action pictures (90%), Spanish speakers used both manner verbs and path verbs (46% and 38%, respectively) in their descriptions.

In this paper, we consider the extent to which cross-linguistic differences in how manner of motion is typically lexicalized may affect peoples' behavior on nonlinguistic similarity judgment tasks. The majority of work concerned with showing how differences in cross-linguistic usage covary (or not) with differences in performance on nonlinguistic tasks has been in the area of color perception (Berlin & Kay, 1969; Rosch 1973), spatial cognition (Brown & Levinson, 1993; Levinson, 1996; Pederson, 1995; Pederson, Danziger, Wilkins, Levinson, Kita, & Senft, 1998), and object categorization (Malt, Sloman, Gennari, Shi, & Wang, 1999). There has been relatively little work on how cross-linguistic differences in verb-conflation tendencies affect performance on nonlinguistic tasks. The experimental work focusing on crosslinguistic differences in how motion events are encoded has almost always involved overt language use. For instance, some research has compared how speakers of different languages describe particular motion events (Naigles, *et. al.,* 1998; Slobin, 1996), and one study (Naigles & Terrazas, 1998) looked at what component of motion (manner or path) speakers of different languages expected novel verbs to encode. In this latter study, Naigles and Terrazas had participants learn novel verbs for different actions in two different sentence frames: a manner frame ("She's kradding toward the tree.") or a path frame ("She's kradding the tree").

Participants were then shown two variants of the original action: one that retained the original manner of motion (manner screen) and one that retained the original path of motion (path screen). Naigles and Terrazas found that native speakers of Spanish were more likely to choose the path variants when told to "point to *kradding*" and that native English speakers were more likely to choose the manner variants. These findings clearly show that one's language plays a role in the interpretation of novel words, but they are inconclusive with respect to whether or not language may influence nonlinguistic cognitive processes. It would be striking to find language effects on nonlinguistic processes and conceptualizations as exhibited in tasks in which language is not engaged. The following experiments were designed to do just this by testing speakers of languages that differ in the extent to which manner is encoded within verbs: English (which relatively often conflates manner and motion) and Spanish and Japanese speakers (which less often conflate manner and motion) (Talmy, 1985, 2000). If an individual's performance on a nonlinguistic task is consistent with the verb-conflation patterns of his/her native language, then based on previous experiments using overt language use (Naigles & Terrazas, 1998; Naigles *et. al.*, 1998; Slobin & Hoiting, 1994), we would expect clear preferences by English speakers to attend to the manner components of motion events. This work would also lead us to expect Spanish speakers to exhibit a weak preference to attend to the path components of motion events. Because of the paucity of similar research on Japanese, our predictions are based upon how Japanese is classified typologically, which suggests that their preferences should exhibit a pattern close to those of Spanish speakers.

EXPERIMENT 1

The design of Experiment 1 was adapted from Naigles and Terrazas' study (1998), which used four short video clips of people performing a set of familiar actions (e.g., skipping, crawling). In our experiment, we developed 10 novel animations that contained motions that were not easily labeled with a single lexical item. These animations consisted of a single object moving in a novel manner (achieved, for example, by having the object simultaneously rotating or oscillating on two different axes) toward, and eventually through, a goal, such as a window- or door-like "exit" from the scene. For each of the 10 novel animations, two variants were created, one that retained the manner but not the direction of motion and another with retained the direction but not the manner of motion. The animations were presented one after the other: an initial "study" screen containing one animation that was immediately followed by the two variants, which were displayed side-by-side. Participants were asked to indicate which of the two animations (*path* or *manner*) was most similar to the first animation of the animation set (see Appendix I for the complete instructions). See Fig. 1 for an example.

Method

Participants

Four groups of participants were tested. All subjects were asked to read and sign a consent form; all rights of subjects were protected. Twenty-one monolingual English speakers, undergraduates at the University of Arizona, participated for course credit. Twenty-three native Spanish speakers participated. These participants were all graduate students in the Spanish Department at the University of Arizona and were paid for their participation. Forty-one native Japanese speakers participated. They were also paid for their participation. Seventeen were Japanese-English bilingual graduate

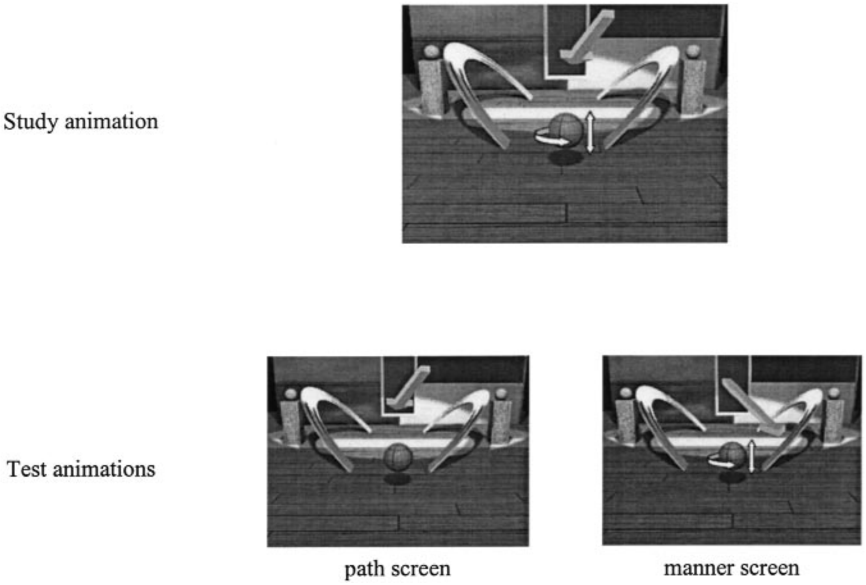

Fig. 1. In Experiment 1 the study animation was viewed separately from, and before, the test animations. The test animations were viewed simultaneously. The “path screen” retains the path of motion from the study animation. The “manner screen” retains the manner of motion from the study animation. Participants were asked to select one of the test animations as being most similar to the study animation. In Experiment 2, all three animations were viewed simultaneously and participants were asked to indicate which animation on the bottom was most similr to the one on the top.

students at the University of Arizona and 24 were monolingual undergraduates at Niigata Sangyo University and Niigata Institute of Technology in Niigata Prefecture, Japan.

Materials

Ten 40-second animations were created using 3D Studio Max™ software. Crucial to the creation of these animations was the novelty of each object's manner of motion so as to prevent participants from easily labeling the motions. It was possible to refer to some of the motions using familiar verbs in phrases (e.g., "jumping and spinning"), but none of the motions could be referred to with a single lexical item in English, Spanish, or Japanese. In addition to the 10 original animations, two variations (*path* and *manner*) of each original were created for the similarity judgement task. These were identical to the originals in every way except for the direction of object traversal (the *manner* screen, in which manner of motion was preserved) or the axis on which the object rotated or oscillated (the *path* screen). Previous research has shown that for a path interpretation of a motion scene to occur at all, the object's path must be made salient (Naigles & Terrazas, 1998; Naigles *et al.*, 1998). This was achieved in the study animation by having the moving object in each animation traverse a perceptual boundary within the scene and exit the scene just before the animation's conclusion. The animations were presented on a computer monitor directly in front of the participant.

Procedure

Each participant was tested individually in a laboratory test booth with dimmed lighting. The participant sat in front of the computer monitor, while the experimenter sat behind the participant. Instructions for the experiment were presented on the computer monitor in English, Spanish, or Japanese, depending on the native language of the participant. Participants were asked to pay close attention to each original animation and then to point to one of the subsequent pair of animations that they thought was more similar to the original animation. The order of each set of animations was presented pseudo-randomly across the experiment by creating four different lists. The side on which the manner and path screens appeared was counterbalanced across lists. For each participant, path and manner screens appeared equally often on the left and right sides.

Results and Discussion

The results appear in Table I. As is clear from the table, participants' choices in this nonlinguistic task were consistent with the verb-conflation

Table I. Participants' Mean Number of Manner Screens Chosen, Experiment 1

	Mean number of manner screens chosen
English monolinguals ($n = 21$)	8.8
Spanish-English bilinguals ($n = 23$)	5.6
Japanese monolinguals ($n = 24$)	6.0
Japanese-English bilinguals ($n = 17$)	6.0

tendencies of their native language: the native English speakers selected the manner screen significantly more often than the Spanish and Japanese speakers. Separate analyses of variance (ANOVA) with participants (F1) and items (F2) as random variables revealed a significant group effect [$F_1(2,82) = 9.527$, $p < .001$; F_2 $(2,19) = 35.838$, $p < .0001$].[4] Planned comparisons showed that English speakers chose manner screens significantly more often than Spanish speakers [$F_1(1,42) = 12.55$, $p < .001$; $F_2(1,9) = 67.54$, $p < .0001$] and more often than Japanese speakers [$F_1(1,60) = 26.57$, $p < .001$; $F_2(1,9) = 62.09$, $p < .0001$]. Further, the Japanese speakers chose manner screens significantly more often than chance [$F(1,40) = 6.569$, $p < .05$], whereas the Spanish speakers showed no preference. This latter difference is difficult to interpret. It is possible that Japanese speakers use manner verbs more frequently than Spanish speakers. Or it may be just an issue of statistical power (41 Japanese-speaking participants versus 23 Spanish-speaking participants).

The results of the present experiment indicate that individuals with different linguistic systems attend to different aspects of events in making their similarity judgments, raising a question about the underlying psychological mechanism responsible for the observed differences. There are several possibilities. One is that the character of the language affects the perceptual system; that is, people actually see differently as a result of the characteristics of their native language. Another is that language differences produce differences in the aspects of a scene to which people attend. A third possibility is that the participants were covertly using language to be able to better remember the "study" scene. This possibility suggests (quite plausibly) that the ease with which the different aspects of motion may be expressed in a particular language affects the language representation that speakers create: English speakers would more readily formulate a verb phrase containing manner information than either the Spanish or Japanese speakers because they do so more frequently. Experiment 2 was designed to explore these possibilities.

---

[4] The monolingual and bilingual native Japanese speakers performed identically, so the data from these two groups were combined for all subsequent analyses.

EXPERIMENT 2

In this experiment, we tested two groups of native English-speaking participants using exactly the same stimuli as in Experiment 1. One group was tested using the same procedure as in Experiment 1, except that they were tested as a group in a classroom. Another group was also group-tested in a classroom, but with a modified procedure: the “study” animation and the two similarity-judgment variants were all presented simultaneously, so there was no memory component. We restricted this study to native English speakers because, of the three groups tested in the first experiment, the native English speakers showed the greatest preference for the manner frames. If we assume that the memory component of the task elicited the use of language, which in turn caused the English speakers to make so many “manner” selections, then eliminating the memory component should eliminate this effect.

Method

Subjects

Sixty-three undergraduate monolingual English-speakers at the University of Arizona participated for course credit. Twenty-four participated in the nonmemory version of the task, and 39 participated in the memory version (the replication).

Materials and Procedure

The stimuli were identical to those in Experiment 1. For the nonmemory group of participants, the original animation of each animation set was presented simultaneously with its path and manner variants. The original animation was located at the top of the screen in the center, and the path and manner variants appeared side by side, below the original animation. The path and manner variants appeared equally on both left and right sides of the screen. For the replication group, the study animation was presented first, followed by the pair of test animations.

All participants were tested in a single session in a classroom with the lights dimmed. An LCD projector was used to present the animation sets on a white screen in front of the classroom. Participants were separated to prevent copying or “working together.” Participants in the nonmemory condition were provided with answer forms with 10 sets of squares arranged in the same triangular configuration as the animations on the screen. They were instructed to “circle the movie on the bottom that is most similar to the one on the top.” The participants in the memory condition indicated their screen choices on a scantron answer sheet.

Results and Discussion

The mean number of “manner” selections appear in Table II.

The nature of the task has a patent effect on the number of manner screens that native English speakers select: When the task includes a memory component, manner frames are selected more than 80% of the time; when the task makes no memory demands, the selection of manner frames drops to close to 50%, an incidence that does not differ significantly from the number chosen by native Spanish and Japanese speakers in Experiment 1 [$F_1(2,85) < 1$; $F_2(2,18) < 1$], but does significantly differ from the number chosen by native English speakers in both Experiment 1 [$F_1(1,43) = 19.91$, $p < .0001$; $F_2(1,9) = 13.15$, $p = 0.006$] and Experiment 2 [$F_1(1,61) = 27.70$, $p < .0001$; $F_2(1,9) = 10.89$, $p = .009$]. Because the nonmemory task in Experiment 2 required close perceptual analysis of the stimuli, and because there was no language effect, we can rule out the possibility that English speakers *see* motion events differently or pay more attention to the manner of motion. Rather, the conclusion seems inescapable that the apparently nonlinguistic task used in Experiment 1 actually encouraged participants to encode the scenes linguistically. Given this analysis, it is hardly surprising that participants’ choices were consistent with the verb-conflation patterns of their native language.

General Discussion

This study used a similarity-judgment task to investigate the degree to which cross-linguistic differences in verb-conflation tendencies affected participants’ conceptualizations of motion events. Experiment 1 showed that when people are asked to remember a motion event for even a very short period of time, the components of the event that they encode into memory (MANNER or PATH) are congruent with the verb-conflation patterns of their native language. We also found that this language effect was task dependent. In Experiment 2, we changed the procedure so that the “study” animation and the two similarity-judgment variants were all presented simultaneously. Under these conditions, we found that the language effect observed in Experiment 1 disappeared, suggesting that in Experiment 1, participants had been covertly using language to enhance their memory of the study scene.

Table II. English Monolingual Participants’ Mean Number of Manner Screens Chosen, Experiment 2

	Mean number of manner screens chosen
Non-memory condition ($n = 24$)	5.1
Memory condition ($n = 39$)	8.5

These experiments clearly bear on the interpretation of linguistic relativity effects. Typically, work in this area is correlational in nature: a particular cross-linguistic difference is correlated with behavioral differences among individuals from different linguistic communities. The studies rarely report conditions under which the observed linguistic-relativity effect disappears (although see Kay & Kempton, 1984, for a counterexample), leaving one to speculate on the locus of interaction between language and behavior. The experiments reported here reveal a task-dependent effect that strongly suggests that when a task requires participants to hold information in mind, they recruit language to rehearse the information.

It has been known for quite some time that people are much better in memory tests when they can use language. For example, in a seminal experiment, Bower, Karlin, and Dueck (1975) showed that participants' recall of nonsensical "droodles" was positively affected when verbal cues were given during memory encoding. Line drawings depicting, for example, what looked like two bird legs protruding from a hole were indecipherable until a cue such as "an early bird got a very strong worm" was given. When given verbal cues, participants were significantly better at recalling the pictures in the experiment than when there were no verbal cues. Bower *et al.* (1975) argued that the verbal cues "aroused appropriate schemata" into which the drawings could be placed—that is, the verbal cues provided a context for interpreting the droodle—which helped participants interpret the pictures and thus improved memory for them.

It makes sense that people would rely on a "linguistic encoding" strategy when trying to remember stimuli for which it is easy to generate linguistic descriptions. In a recent study, Zelinsky and Murphy (2000) showed that people make use of verbal labels to enhance their memory encoding of familiar and easy-to-label objects when performing a short-term recognition task. Zelinsky and Murphy presented participants with an array of four pictured objects, followed very shortly afterward (2.5 s) by a probe picture of a single object that participants had to indicate as being part of the original set or not. They found that when participants were scanning the array, they looked longer at pictures of objects with longer labels (e.g., "envelope" versus "cup"), suggesting that in order to more readily perform the task, the participants encoded the verbal labels of the objects into working memory. When presentation conditions were manipulated so that participants did not have to remember the objects, they did not manifest this pattern of looking times, suggesting that they were not generating the labels of the objects.

It is worth highlighting that although the Zelinsky and Murphy (2000) study differed from our own in that their task was explicitly a memory task and their picture stimuli were easy to label, their results converge with ours.

The findings of the present study have important implications for how much of the work reporting linguistic-relativity effects should be interpreted and how future studies investigating questions of linguistic relativity are designed. Some of the most compelling evidence for linguistic relativity effects has come from studies in which an effect is observed in participants' performance on a presumed nonlinguistic task. Typically, however, these studies required participants to encode some set of stimuli into working memory in order to perform the task given to them. For example, a task used by a number of investigators (e.g., Brown & Levinson, 1993; Levinson, 1996; Pederson, 1995; Pederson, *et al.*, 1998) requires participants to remember an array of toy animals in front of them, which are all facing in the same direction. The participant is then rotated 180 degrees and is asked to recreate the array. The variable of interest is the direction that the animals face in the reconstructed array. If the participant initially conceptualizes the animals facing to his/her right (i.e., *relative frame of reference*), then the participant should reconstruct the array from memory so that the animals face to the participant's right. If, though, the animals in the original array are conceptualized as facing north (i.e., *absolute frame of reference*), then the participant should reconstruct the array so that the animals face north. Results from several of these types of studies (Brown & Levinson, 1993; Pederson, 1995; Levinson, 1996; Pederson *et al.*, 1998) have shown that an individual's conceptualization of spatial relations is strongly correlated with the way that individual's linguistic system encodes spatial frames of reference. That is, participants from language communities in which an absolute frame of reference is preferred in language use tend to perform this nonlinguistic spatial task using an absolute frame of reference, and participants from language communities in which a relative frame of reference is employed in language use tend to perform this same task using a relative frame of reference. The results from the present study suggest that such tasks are probably not entirely nonlinguistic. Given the possibility that participants invoke language when they have to remember information for even a brief period (such as the time it takes to turn around and recreate an array), it seems eminently plausible that findings of language affecting behavior arise in just those cases when language is used, even covertly.

APPENDIX I

Instructions were as follows:

In this experiment, you will be asked to group different "movies" together according to how similar they are. The movies are of objects moving through an artificial environment. You will first see a single film clip. Please pay close attention to this film clip, and try to remember the details. When the first clip is finished playing, it will disappear and you will see two more clips. Your task is to say which of those two film clips is most similar to the previous one.

REFERENCES

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

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition, 3*(2), 216–220.

Brown, P., & Levinson, S. C. (1993). Explorations in Mayan cognition. Working paper no. 24, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.

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

Kageyama, T. (1996). *Dooshi-imiron; gengo to ninchi no setten* [Verb-semantics: the point of contact for language and cognition] [in Japanese]. Tokyo: Kuroshio.

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

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Cross-linguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (109–169). Cambridge, MA: MIT Press.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory & Language, 40*(2), 230–262.

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

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

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

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language, 74,* 557–587.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350.

Slobin, D. I. (1996). Two ways to travel: Verbs of motion in English and Spanish. In M. Shibitani, & S. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (195–219). Oxford, England: Oxford University Press.

Slobin, D. I., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society* (487–505). Berkeley, CA: Berkeley Linguistics Society.

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

Talmy, L. (2000). *Toward a cognitive semantics, Vol. II: Typology and process in concept structuring.* Cambridge, MA: MIT Press.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science, 11,* 125–131.