Spatial Position in Language and Visual Memory:
A Cross-Linguistic Comparison

Solveig Bosse (solveig@udel.edu)
Department of Linguistics & Cognitive Science, 46 E. Delaware Ave
Newark, DE 19716 USA

Anna Papafragou (annap@udel.edu)
Department of Psychology, 109 Wolf Hall
Newark, DE 19716 USA

Abstract

German and English speakers employ different strategies to encode static spatial scenes involving the axial position (standing vs. lying) of an inanimate figure object with respect to a ground object. In a series of three experiments, we show that this linguistic difference is not reflected in native speakers' ability to detect changes in axial position in non-linguistic memory tasks. Furthermore, even when participants are required to use language to encode a spatial scene, they do not rely on language during a recognition memory task. These results have implications for the relationship between language and visual memory.

Keywords: Positional verbs, visual memory, space, language and thought

Introduction

It has often been observed that languages make available different strategies to encode spatial relations (Ameka & Levinson, 2007). A question of central interest within the cognitive sciences is how these cross-linguistic differences interact with underlying spatial representations recruited in spatial memory and other cognitive processes. According to an influential position, language exerts a strong influence on cognitive processes (Levinson, Kita, Haun & Rasch, 2002). Based on several cross-linguistic experiments involving spatial tasks, Levinson et al. proposed that spatial frames of reference provided by people's native language affect how people remember spatial arrays. Crucially, such linguistic effects are argued to emerge even when no overt linguistic labels accompany encoding of the spatial scene. The idea is that obligatory spatial distinctions made within one's native language direct attention to those aspects of spatial representation - thereby affecting spatial memory.

According to a different position, effects of native language on mental representation and memory are more limited. For instance, studies have shown that, despite differences in encoding motion events between English and Greek, memory for aspects of motion does not differ across speakers of these languages (Papafragou, Massey & Gleitman, 2002; cf. Gennari, Sloman, Malt & Fitch, 2002 for related results on English vs. Spanish). Other work has also suggested an independence of memory from cross-linguistic differences in spatial encoding (see Munich, Landau & Dosher, 2001; cf. reviews in Gentner & Goldin-Meadow, 2003).

The question of whether (and how) cross-linguistic differences might affect memory for spatial scenes is related to the question of whether (and how) the explicit presence of linguistic labels during spatial encoding might affect memory. Effects of overt labeling, even though not as deep and pervasive as the effects proposed by Levinson et al. (2002), are still important for understanding how language interfaces with other cognitive faculties. Several studies have shown that explicit spatial language can affect spatial memory – even though the scope and mechanisms responsible for such effects are still not well understood. For instance, there is evidence that memory for motion events can be biased depending on whether path (*exit*) or manner (*skip*) verbs accompany the events, regardless of whether the verbs are provided by the experimenter (Billman & Krych, 1998) or generated by participants (Billman, Swilley & Krych, 2000). Relatedly, Feist and Gentner (2007) showed that providing participants with spatial language can influence their behavior in a recognition task. Specifically, viewing ambiguous spatial representations paired with spatial prepositions (e.g., *on*) resulted in a false memory bias towards typical portrayals of the relation encoded by the prepositions. In another demonstration, Archambault, O'Donnell and Schyns (1999) showed that the level at which an object is categorized (general, e.g. "a mug", or specific, e.g. "Steve's mug") influences the time it takes people to detect a change in a picture containing the object. If objects are known at a specific (individual) level, then the changes are detected faster than if the objects are known on a general level. Crucially, in this study, the level of categorization was provided through linguistic labels prior to the main testing phase.

In this paper, we explored the extent of the influence that language can have on spatial memory (including contexts with and without overt linguistic encoding). We focused on an area of spatial encoding that has only recently begun to receive attention in the literature – namely, positional systems (see Ameka & Levinson, 2007) – and compared two languages, English and German, that differ in a specific aspect of spatial-positional encoding. Specifically, German naturally uses positional verbs that specify the axial orientation of the figure object: an object that is perceived to

be upright (whose vertical height exceeds its width) is described with the verb *stehen* 'stand', while an object whose horizontal width exceeds its vertical height is described with the verb *liegen* 'lie'. Although English has equivalent verbs and uses them for humans and other animates, the positions of inanimate objects are typically and canonically described with the English copula *be* (Kutscher & Schultze-Berndt, 2007). Consider, for example, the two scenes in Figure 1. In German, the two scenes would be canonically described with two difference sentences:

1) Das Buch **steht** auf dem Stuhl. (Fig.1a)
   the book stands on the chair
2) Das Buch **liegt** auf dem Stuhl. (Fig.1b)
   the book lies on the chair

In English, however, both scenes can canonically be described with the same sentence:

3) The book **is** on the chair.

Figure 1a: *stehen* 'stand'   Figure 1b: *liegen* 'lie'

Since this aspect of linguistic encoding represents an obligatory way of encoding spatial position in German but is absent from the corresponding English sentences, we considered it a particularly appropriate test case for the hypothesis that linguistic distinctions influence non-linguistic memory. In a series of experiments (Exp. 1, 2 and 3), we investigated whether this difference in linguistic encoding is mirrored in performance in a (nonverbal) memory task. If language influences memory, then German speakers should perform better than English speakers on a recognition memory task involving changes of posture such as those between Figure 1a and 1b, *even when no language is overtly present as spatial scenes are committed to memory*. If language does not influence memory, native speakers of German and English should perform similarly on the recognition task.

Our studies also addressed the question whether the overt presence of linguistic labels during the encoding of spatial scenes affects memory performance (Exp. 3). Again if overt language affects memory for spatial scenes, German speakers should have an advantage in recognition memory

targeting axial position of a figure object compared to English speakers. If recognition memory is independent of overt labeling, no difference in memory for positionals should exist between English and German speakers.

Experiment 1

Experiment 1 was conducted to collect linguistic data to confirm the difference in the use of positional verbs between English and German. The experiment also tested memory performance for the corresponding positions after participants had freely inspected a set of spatial scenes.

Participants

Twenty-six native speakers of German and 28 native speakers of English participated. The German speakers were recruited at Ruprecht-Karls-Universität Heidelberg in Germany, and the English speakers at the University of Delaware in the U.S. None of the German speakers spoke English natively, although almost all had studied it in school, usually alongside other languages. Similarly, none of the native English speakers had native speaker proficiency in German. Equal numbers of men and women were included.

Stimuli

The stimuli consisted of 40 pairs of pictures, taken in color with a digital camera. Each picture depicted two everyday household objects arranged in a particular way. The objects were placed in mostly unconventional pairings (e.g., a boot with a frying pan, a teabag with a wine glass) so that the participants would not focus on the position of the objects but rather on their unpredictable combinations of the objects. Each object appeared in one and only one pair of pictures.

Sixteen of the 40 pairs were test items, which always displayed a figure object on top of a ground object. One picture in each pair depicted the figure object in a standing, vertical position, consistent with the German verb *stehen*, while the other picture depicted the figure object in a lying, horizontal position, consistent with the German verb *liegen*. The position of the ground object was the same in both pictures (see Figure 1 for an actual example drawn from our stimuli). The figure objects had to be medium-sized items that balanced well, could be placed in either a standing or lying position, and would look acceptable in both. We avoided objects that resembled animate beings (e.g., dolls) because English uses *stand* and *lie* for human beings in the upright or horizontal position. In fact, most everyday objects have an inherent orientation — they either stand up or lie flat in their natural state. Therefore, we supplemented our small number of orientation-free figure objects (e.g., lipstick, a roll of paper towels) with an equal number of figures that either naturally "stand" (e.g., a wicker basket) or naturally "lie" (e.g., a wallet), in order to avoid any bias created by unusual positioning.

Another 8 of the 40 pairs of pictures were changing control items (i.e., they involved changes that were

unrelated to the stand-lie distinction). In the changing control pictures, the two objects were placed in a non-support relationship in at least one of the two pictures. Such relationships involved attachment (e.g., a paper clip on a pen), containment (e.g., a banana in a bowl), or piercing (e.g., a knife in an apple). The difference between the two arrays in each pair were either changes of state (e.g., a banana becomes a peeled banana) or non-axial changes in position (e.g., a paper clip originally attached to the cap of the pen becomes attached to the body of the pen).

Finally, 16 of the 40 pairs of pictures were non-changing control items. The two members of each pair were identical to each other and depicted relationships of support (with one object resting on top of another), attachment, or containment.

These pairs of pictures were arranged for display in two lists of 40 pictures each. One picture of each pair became part of List 1 and the second picture of each pair became part of List 2. Within each List, half of the test items depicted a standing relation and half a lying relation. Lists 1 and 2 displayed pictures in two different random orders. We also created two more lists (Lists 3 and 4) by reversing the presentation order of Lists 1 and 2. For the memory task, we arranged these lists into four different working orders that varied in terms of which list was used during the initial (encoding) phase vs. the second (memory) phase (List 1 vs. 2, 2 vs. 1, 3 vs. 4, or 4 vs. 3 respectively).

For the language task, we selected a subset of these stimuli for presentation. Specifically, we only used List 1 and List 2 but omitted the non-changing control items such that each list contained 16 test and 8 changing control items only.

Procedure

Language Task For the language task, we tested 10 German speakers and 12 English speakers. Participants viewed either the (shortened) List 1 or the (shortened) List 2. They were told that each picture would depict two household objects paired together, and were asked to describe each arrangement with a single complete sentence. Participants recorded their responses on a lined answer sheet and controlled the pace of the task by advancing the display themselves.

Memory Task For the memory task, we tested 16 German speakers and 16 English speakers. None of these had participated in the language task. Each participant was assigned to one of the four stimuli orders. The participants were simply told that they would see a set of pictures and their task was to look at the pictures carefully. During this encoding phase, each picture appeared for two seconds before the display automatically advanced to the next picture. Then participants were told that they would view a second set of pictures and were instructed to verbally provide fast judgments of whether each picture was the "same" or "different" (i.e., whether the exact same picture had appeared in the first round, or the picture was similar to a picture that had appeared before but was also recognizably

changed). The pictures in the memory phase were also displayed for two seconds each. If a participant did not provide an answer within those two seconds, his or her response was discarded.

Results and Discussion

Language Task As the dependent variable, we calculated the percentage of answers that included a positional term for each language group. All positional information was encoded in verbs, namely German *stehen* and *liegen* and their English equivalents *stand* and *lie*. German speakers encoded position 90% of the time while English speakers encoded position only 32.3% of the time. This difference is significant (two-tailed t-test, p<.001). Thus, as expected, native speakers of German are more likely to encode the detailed spatial position of a figure object than English speakers.

Memory Task The results for this task are displayed in Figure 2. (All error bars in this paper indicate standard error.) For this and the following memory experiments, the dependent variable is the percentage of correctly identified pictures. An ANOVA with Language (German, English) and Trial (Test, Changing Control, Non-Changing Control) as factors returned only a main effect of Trial (F(2,29)=22.04, p<.0001). The effect is due to a significant difference between Test items (M = 69.73) and Changing Control items (M = 88.49; p<.0001), as well as a difference between Test items and Non-Changing Controls (M = 91.51; p<.0001). Thus, despite differences between English and German in the labeling of spatial position, English speakers did not perform differently from German speakers in memory for spatial position.
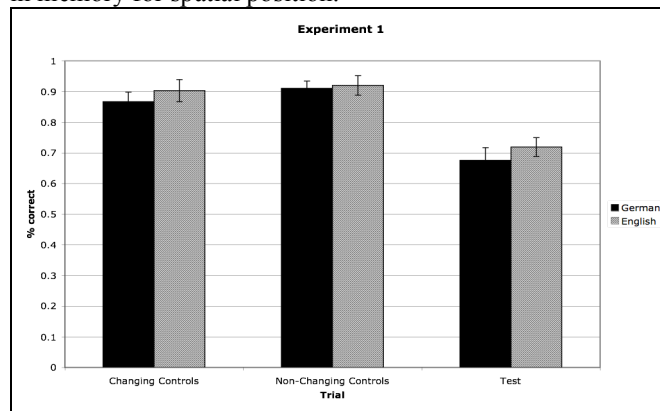

Figure 2: Accuracy in Memory Task of Experiment 1

Experiment 2

One possible explanation for the lack of native language influence in the memory task of Experiment 1 is that participants were not warned that memory for pictures would be tested. It is possible that prior knowledge of the nature of the task would make people more likely to recruit linguistic resources to encode the objects and relations in the pictures in anticipation of later testing. In Experiment 2, we

tested this hypothesis. Specifically, we replicated Experiment 1 but made participants aware of the fact that they would have to remember the pictures. To further bolster the opportunity to use linguistic labels (and store both the labels and the visual scene in memory) we introduced a temporal gap between pictures during the encoding phase. We reasoned that this lag of time would allow participants to encode stimuli verbally even if they were not specifically instructed to do so.

Participants

Sixteen native speakers of German were recruited at the Carl-von-Ossietzky Universität Oldenburg in Germany, and 16 English speakers were recruited at the University of Delaware in the U.S. None of the German speakers spoke English natively, although almost all had studied it in school. Similarly, none of the native English speakers had near-native speaker proficiency in German. Approximately equal numbers of men and women were included. None of these people had participated in Experiment 1.

Stimuli

The same materials as in Experiment 2 were used.

Procedure

The same procedure as for the memory task in Experiment 1 was used but with two modifications. First, participants were told that this would be a memory experiment and that they needed to remember the pictures they would see for a later recognition test. Second, 3 seconds of blank screen were inserted between pictures in the encoding phase.

Results and Discussion

The results are displayed in Figure 3. An ANOVA with Language and Trial as factors returned only a main effect of Trial $(F(2,29)=26.42, p<.0001)$. This effect is driven by lower performance on Test items (M = 71.29) compared to the Changing Controls (M = 90.23; $p<.0001$) and the Non-Changing controls (M = 93.16; $p<.0001$). Thus even when participants know that they are participating in a memory experiment and are given the opportunity to encode the stimuli linguistically, linguistic encoding does not appear to affect the outcome of recognition memory.
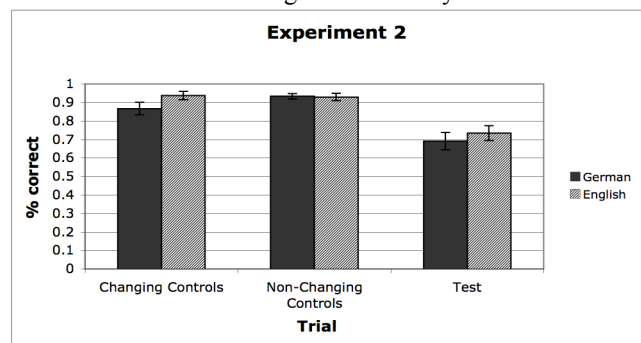

Figure 3: Accuracy in Memory Task of Experiment 2

Experiment 3

Participants in Experiment 2, even though given the opportunity to encode the visual scenes linguistically, did not necessarily do so. It is an open question whether, under different conditions (e.g., a more difficult task), participants might spontaneously recruit labels implicitly as an additional encoding strategy (which would lead to English-German differences in memory accuracy here). Experiment 3 followed the basic method of Experiment 2 but introduced a novel manipulation to address this question.

Specifically, we included a Non-Linguistic Shadowing condition in which participants engaged in a secondary task (shadowing a rhythm by tapping) while inspecting the scenes: crucially, this shadowing task did not engage the language faculty. We hypothesized that, because of the high cognitive load imposed by the secondary task, participants would be likely to recruit language as an additional means of encoding the scenes in preparation for the memory test. If so there could be language-specific patterns in memory performance. For comparison purposes, we also included a Linguistic Shadowing condition in which participants engaged in a comparable secondary task that blocked the language code (verbally shadowing a rhythm). This task was not expected to lead to recruitment of labels during encoding (or to cross-linguistic differences in spatial memory). Hermer-Vazquez et al. (1999) showed that these two types of shadowing tasks impose the same cognitive load but employ different cognitive resources. Thus, labeling could be possible with Non-Linguistic Shadowing but not in Linguistic Shadowing.

Experiment 3 also tested the hypothesis that, when forced to provide linguistic labels explicitly, participants would use such labels later during the recognition phase (thus triggering language-specific effects on memory performance). In a Linguistic Completion condition, participants were asked to fill out a sentence after each scene describing the scene they saw; critically, they had to provide the spatial verb describing the relationship between the figure and ground object. German speakers were expected to produce more positional verbs than English speakers. Importantly, if labels can affect visual memory, we should expect an advantage for German speakers compared to English speakers in later recognition of standing vs. lying object positions. This manipulation provides a powerful test for the hypothesis that labels affect memory performance by virtually guaranteeing the presence of labels (hence of cross-linguistic labeling differences) during the initial inspection of visual scenes.

Participants

Thirty-six native speakers of German were recruited from either the Carl-von-Ossietzky Universität Oldenburg or the Gymnasium Nordenham in Germany. All had learned English but none of them spoke it natively. Thirty-six native speakers of English were recruited at the University of Delaware. No native speaker of English was fluent in

German. None of these participants had participated in Experiment 1 or 2. Approximately equal numbers of men and women participated.

Stimuli

The same materials as in Experiment 2 were used.

Procedure

Participants were randomly assigned to one of three conditions:

Non-Linguistic Shadowing Procedure was as in Experiment 2 but participants wore headphones during the encoding phase and listened to an irregular rhythm. Their task was to repeat the rhythm by tapping it with their fingers on the table.

Linguistic Shadowing Participants followed the same procedure as those in the Non-Linguistic Shadowing condition except that they had to repeat the rhythm constantly using the syllable "na" (they had to say the syllable loud enough for the experimenter to hear it).

Linguistic Completion Procedure was as in Experiment 2 with some modifications. After each picture in the encoding phase, instead of a blank screen, participants saw a screen displaying a sentence. The sentence was presented for 3 seconds and appeared in the native language of each participant. The sentence described the preceding spatial scene but was missing the verb and the ground object. For instance, for Figure 1a above, English speakers saw "The book ____ on the ____." Participants were instructed to read the sentence out loud during the time it was displayed adding in the missing words (the ground object was omitted so that English speakers would not simply have to provide the copula *is* throughout). Sentences were recorded and later transcribed for coding.

Results and Discussion

Non-Linguistic and Linguistic Shadowing Conditions The results from the memory task for these two conditions are presented in Figures 4a-b. For the Non-Linguistic Shadowing condition, an ANOVA with Language and Trial as factors returned only a main effect of Trial ($F_{(2,21)}$ = 10.5, p<.001). The effect is driven by lower performance on Test items (M = 62.6) compared to Changing Controls (M = 83.9) and to Non-Changing Controls (M = 80.2, p<.05). A similar ANOVA for the Linguistic Shadowing condition gave similar results (main effect of Trial, $F_{(2,21)}$ = 13.47, p< .0001, with lower performance on Test items (M = 61.1) compared to Changing and Non-Changing Controls (M = 79.7 and 78.6 respectively, p<.05)). No difference was observed between performance in the two shadowing conditions (p>.05). Thus even in a task with higher cognitive demands that allows for the use of the linguistic code, language does not seem to have an effect on scene representations recovered from memory.

Linguistic Completion As expected, participants' linguistic productions confirmed the asymmetry between English and German: German speakers offered verbs

encoding the (correct) position of the figure object for 73.3% of the Test items; English speakers did so for only 2.8% of these items. This difference is significant (two-tailed t-test, p<.05).
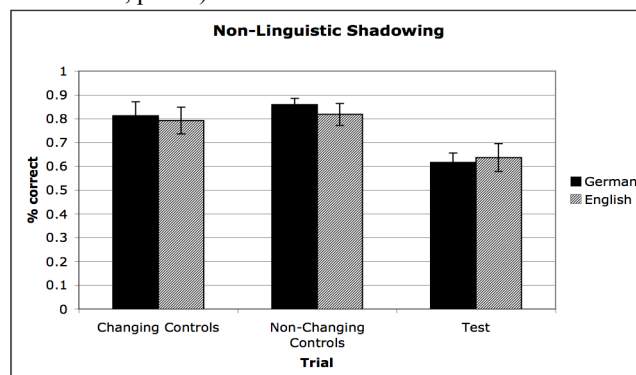

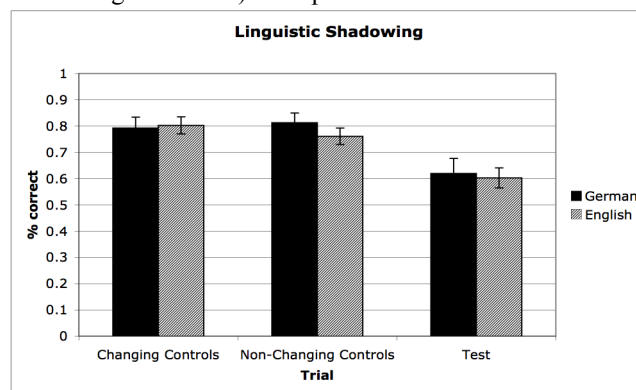Figure 4a: Accuracy in Memory Task (Non-Linguistic Shadowing Condition) of Experiment 3

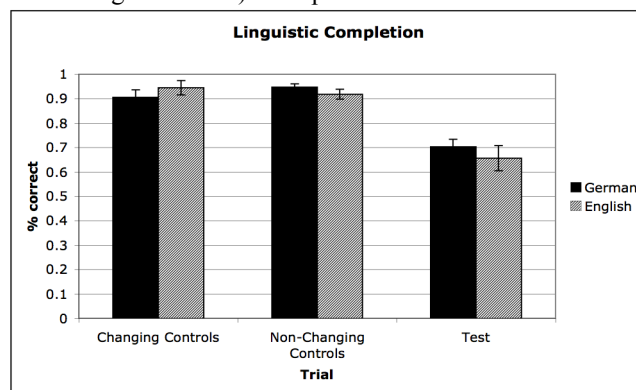Figure 4b: Accuracy in Memory Task (Linguistic Shadowing Condition) of Experiment 3

Figure 4c: Accuracy in Memory Task (Linguistic Completion Condition) of Experiment 3

For the memory data (Figure 4c), an ANOVA with Language and Trial returned only a main effect of Trial ($F_{(2,21)}$=47.29, p<.0001). This effect is driven by lower performance on Test items (M = 66.3) than Changing Controls (M = 92.1) and Non-Changing Controls (M = 93.7). German speakers – unlike English speakers - overwhelmingly mentioned the axial position of the object

in filling out the target sentences but this linguistic encoding did not lead to an advantage in remembering axial position.

General Discussion

In this study, we asked whether differences in the way English and German encode the axial position of a figure object affect recognition memory for axial position. Our results suggest that cross-linguistic differences in positional encoding have no influence on memory for spatial scenes. Specifically, in a variety of contexts allowing or encouraging the choice to encode the scenes linguistically, participants did not appear to make this choice. These results argue against theoretical positions according to which obligatory lexical or grammatical distinctions in a language create cognitive biases in speakers even in situations where no language is overtly present (e.g., Levinson et al., 2002). Our data are consistent with prior finding showing that spatial memory is independent from cross-linguistic differences in spatial vocabulary (Papafragou et al., 2002; Gennari et al., 2002).

A particularly noteworthy aspect of our data is that native language distinctions failed to affect recognition memory even when participants explicitly provided linguistic encoding of the spatial scenes (Linguistic Completion condition of Exp. 3). This finding differs from previous reports which found effects of explicit labeling on visual memory in speakers of a single language (see Introduction). To reconcile these divergent findings, one possibility is that language effects are more likely to surface when labels occur before (as in Feist & Gentner, 2007; Archambault et al., 1999; Billman et al., 2000) or during (as in Billman & Krych, 1998) the encoding of visual scenes rather than after visual encoding has occurred (as in our Exp. 3). In support of this possibility, work by McCloskey and Zaragoza (1985) showed that verbally presented misinformation about an object *after* an object had been viewed (e.g., referring to a hammer as a screwdriver) did not impair participants' ability to later recognize the object, as opposed to a new, previously unmentioned, object. Nevertheless, this explanation cannot account for other work showing that, even when linguistic labels are generated as spatial scenes or events are viewed, they do not necessarily alter visual memory (Papafragou et al., 2002; Gennari et al., 2002).

Another possibility is that language effects are more likely to emerge when the visual scenes to be remembered are ambiguous (Feist & Gentner, 2007) or can be categorized on several levels (Archambault et al., 1999), and thus allow language to play a disambiguating role. Regardless of the specific explanation that will turn out to be correct, the fact that linguistic labels in the Linguistic Completion condition degraded faster than the visual memory of the scenes provides evidence that linguistic and visual representation of spatial position belong to different levels of representation and are potentially independent of each other. Further work is needed to specify the precise factors that affect language intrusions into non-linguistic cognitive processes. Nevertheless, results from the Linguistic Completion task suggest that such intrusions depend on subtle features of the task at hand and do not generalize across all contexts in which language is used to label spatial scenes.

References

Ameka, F. & Levinson, S.C. (2007). Introduction: the typology and semantics of locative predicates: posturals, positionals, and other beasts. *Linguistics* 45, 847-871.

Archambault, A., O'Donnell, C. & Schyns, P. (1999). Blind to object changes: when learning the same object at different levels of categorization modifies its perception. *Psychological Science* 10 (3), 249-255.

Billman, D. & Krych, M. (1998). Path and manner verbs in action: effects of "skipping" or "exiting" on event memory. *Proceedings from the 20th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Billman, D., Swilley, A. & Krych, M. (2000). Path and manner priming: verb production and event recognition. *Proceedings from the 22nd Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Feist, M. & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition* 35 (2), 283-296.

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

Gentner, D. & Goldin-Meadow, S., eds. (2003). *Language in mind.* Cambridge, MA: MIT Press.

Hermer-Vazquez, L., Spelke E., Katsnelson, A. (1999). Sources of Flexibility in Human Cognition: Dual-Task Studies of Space and Language. *Cognitive Psychology* 39, 3-36.

Kutscher, S. & Schultze-Berndt, E. (2007). Why a folder lies in the basket although it is not lying: the semantics and use of German positional verbs with inanimate figures. *Linguistics* 45, 983-1028.

Levinson, S., Kita. S., Haun, D. & Rasch, B. (2002). Returning the tables: language affects spatial reasoning. *Cognition* 84, 155-188.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General* 114, 1-16.

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

Papafragou, A., Massey, C. & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition* 84, 189-219.