Practicing Second Language Spatial Prepositions in a Virtual World

Nora Presson, Michael Heilman & Brian MacWhinney

Carnegie Mellon University

Using Virtual Reality to Teach Spatial Prepositions

Adult second language learners face obstacles to learning that arise from a variety of social, cognitive, and neurological pressures (MacWhinney, in press). As a result, they are often unable to attain native level proficiency (e.g., Johnson & Newport, 1989). However, by emphasizing productive learning strategies (O'Malley & Chamot, 1990) and social/cognitive support methods (MacWhinney, in press), adults can overcome many of these difficulties. One potential source of cognitive support involves linking new second language structures to embodied mental representations (Glenberg & Robertson, 2000 give a model of embodied language). Using this strategy, learners can ground the meaning of new words and relations on specific physical activities and spatial locations. In this study, we explore ways in which a computer interface can be used to provide support for the embodied representation of spatial prepositions in Spanish, and the extent to which this support facilitates the learning of these forms by beginning and intermediate students of Spanish.

Recent psycholinguistic work (Coventry & Garrod, 2004) has underscored the ways in which physical and visual experiences shape our representations of spatial language. For example, participants perceive a man as "under" an umbrella more when the umbrella serves a function of blocking rain. If rain is falling at an angle, than participants will endorse the statement *The man is under the umbrella* for an image of a man with an umbrella angled to block the rain. However, if the rain is falling vertically, this same picture will not be endorsed as semantically accurate (Coventry, Prat-Sala, & Richards, 2001). These *affordances* (semantic information about how different objects interact with the world) affect both linguistic interpretations and object perception. There are also crosslinguistic differences in the expression of spatial relations in language (see Coventry & Garrod, 2004 for a review). For example, the relations English speakers describe with different words as *in* and *on* are not differentiated linguistically in Spanish; both are described as *en*. Such vocabulary differences can also interact with affordances, reinforcing the influence of grounding on concrete language.

Based on the relation between spatial language and the visual world, we predicted that providing embodied practice using spatial terms in a second language would improve learning of those terms. To that end, we built a computerized tutor designed to provide this embodied practice by generating instructions for movement of objects around a virtual room. This tutor required interaction with a visual world in order to complete each trial, and therefore should lead to a greater connection of preposition words to the appropriate visual spatial relations.

We tested this prediction by providing beginning Spanish students practice with common spatial prepositions, either with or without this enactive component designed to encourage embodied representations. The practice we provided to students had several qualities designed to improve the effectiveness of perceptual grounding. First, both the enactive and comparison tasks were Spanish-only, because beginning learners of a second language (L2) have dominant native language (L1) representations and L2 word representations are parasitic on L1 words (Kroll & Stewart, 1994). The purpose of removing English was not to prevent any activation of the much more strongly represented English words (which would be close to impossible), but rather to decrease the need to activate those English representations and make the link between perceptual spatial relations and preposition words in Spanish more salient.

Knowing that native speakers represent spatial language in terms of visual experience does not necessarily imply that visual experience will improve second language learning. However, we assumed that the addition of a visual context would provide some level of perceptual grounding of the spatial preposition through linkage to a real spatial perceptual environment. We hypothesized that this linkage would strengthen the learner's representations of the practiced prepositions. In addition to providing perceptual support for the spatial prepositions, practice with the computerized tutor practice provided opportunities to directly enact the meanings of these terms. We predicted that this enactive component would lead to improved performance by requiring deeper processing of prepositions.

Marley, Levin, and Glenberg (2007) have shown that comprehension can be improved by allowing children to either manipulate objects or watch the experimenter manipulate those objects, as compared to ungrounded free-study. Moreover, imagined manipulation achieves the same improvement as real manipulation (Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004). Further, Glenberg, Brown, and Levin (2007) showed that reading comprehension in young children was improved more by acting out the contents of a passage than by re-reading the passage itself. Glenberg, Goldberg, & Zhu (2009) extended the finding that manipulation improves comprehension to a virtual environment on a computer screen, indicating that virtual movements can create the same strengthening of informational representation that comes from real-world object manipulation (a key assumption in our choice of using a computerized interface).

Based on these findings, the current study had two goals: first, to build and examine the feasibility of a tutor that required acting out simple instructions using spatial prepositions. The second goal was to provide an initial demonstration of the efficacy of the tutor by comparing it on measures of student learning to an equally grounded activity that did not require active interaction with the environment. We predicted the following effects:

1. Although all participants will improve in accuracy and speed of responding after practice, adding an enactive component to a visually represented preposition practice task will lead to larger improvements compared to training with a multiple-choice task.

2. Participants will be able to transfer their knowledge to the test type for which they did not receive training. However, this transfer will be greater for training with the tutor than for training with multiple-choice items.

Method

Participants. Participants were 32 (11 female, 21 male) beginning Spanish students recruited through an online bulletin board at Carnegie Mellon University. All participants were native English speakers who had taken a beginning Spanish class in the previous 2 years.

The average participant age was 28.9, and the average age of first exposure to Spanish was 13.9. Participants had taken an average of 3.2 semesters of instructed Spanish prior to the experiment, but modally rated themselves as beginner or advanced beginners and the mean self-rating was 1.71 (where 1 was beginner and 2 was advanced beginner).

Materials. The preposition tutor was composed of sentences describing rooms containing fixed (landmark) and movable (target) objects (see Figure 1 for an example screenshot). The system generates these sentences using a simple system of templates, following work in natural language generation (Reiter and Dale, 1997). The noun phrase descriptions for movable items consist of a definite determiner (*la* or *el*), followed first by the noun for the item (e.g., “dog”) and then by an optional relative clause including a description of a fixed item (e.g., "that is on the box in the kitchen"), when necessary for resolving ambiguity. The descriptions for fixed items consist of a definite determiner, followed by the noun for the item (e.g., “box”), followed by a prepositional phrase denoting the room in which the fixed item appears (e.g., "in the kitchen"), when necessary for resolving ambiguity.

Eight spatial relations were presented in the tutor: *en* (in / on), *enfrente de* (in front of), *a la izquierda de* (to the left of), *a la derecha de* (to the right of), *bajo* (under), *entre* (between), *detrás de* (behind), and *cerca de* (near). Each tutor sentence was constructed from one of two templates taken from an early pilot with two native Spanish speakers. (e.g., “*Pon la esponja que está en la mesa en la caja en la cocina*.” or “*Recoge la esponja que está en la mesa y ponla en la caja en la cocina*.”). Each sentence contained a source fixed object, a movable object, a target fixed object, and two spatial relation terms.

Figure 1. Screenshot of the tutor interface.

Implementation. Even when considering the time necessary to program the mechanisms for automatic generation mechanisms, there is still a time saving over manual authoring of the exercises. This is because our experiment had to include 108 action types in order to systematically sample each source and target preposition during testing. Apart from its important role in the generation of test trials, automatic generation also allows for the easy creation of an unlimited number of practice tasks during training.

Implementing the task generation required some manual effort for encoding knowledge about the items and spatial relations used in the tutor. For each room image, we defined a set of possible fixed locations. These locations were defined by 2-D coordinates and tagged with depth information, so that moved items would only partially occlude the items positioned behind them. For each fixed item, the valid spatial relations (and the corresponding prepositions) were defined by associating polygonal areas to the coordinates for the item's image. These areas were fairly large in order to allow for valid variation in the position at which students choose to drop an item (e.g., for “under the chair”, any position below the seat and between the legs of the chair is valid). Each spatial relationship also had a canonical position. After a student dropped an item, it was always moved to this expected position in order to simplify the implementation and avoid ambiguity and unnecessary item overlap. Also, some rooms include fixed items as part of the background (e.g., a sink and counter in the kitchen room). These are encoded in the same way as the other fixed items.

A few semantic restrictions were also added to the fixed items and rooms. Certain rooms and items were labeled as being "indoor" or "outdoor" items. Also, some fixed items (e.g., paintings) could only appear on room locations tagged as being part of a wall. There were no restrictions on movable items: they could occur in any room or with any fixed item. Extensions to this semantic system might allow for more plausible actions.

The system used simple 2-dimensional graphics to simulate 3-D relationships in spaces. We did not scale images by their depth, because we used a shallow depth of field when displaying the items (i.e., items are all presented as being roughly the same distance from the user's point of view). The system was implemented using Adobe Flash 10.0 and Actionscript 3.0. Therefore, the system can be used over the web on most current personal computers without any special installation or configuration requirements. For experiments, events were logged by sending HTTP requests from Actionscript code in the system to a simple PHP server, which simply wrote text strings from the tutor to a log file.

Procedure. After completing an informed consent form and demographic questionnaire, participants were given a list of all vocabulary used in the experiment (objects and preposition words along with words used in the sentence stimuli). They were not instructed to try to memorize the words, but rather to familiarize themselves and ask any questions about unrecognized words.

Each participant was assigned to one of two between-subjects conditions. Half of the participants trained with the tutor interface, and half of the participants trained with the multiple-choice interface.

At the beginning of a tutor trial, the participant saw a room laid out with fixed objects (landmarks that could not be moved, e.g. a tree or a sink) and movable objects (e.g., a ball or a cup). Spatial relations were represented as the relationship between a fixed object and a movable object. In the tutor interface, each trial represented two spatial relations: the relation between the *source* fixed object and the movable object (e.g., "Pick up *the ball that is under the tree* and put it on the chair."), and the relation between the target fixed object and the movable object (e.g., "Pick up the ball that is under the tree and put it *on the chair.*"). The source spatial relation was visible at the start of the trial. The participants' task was to create the correct target spatial relation by moving the source movable object to the correct target location.

The multiple-choice interface was generated from screen-captures of the starting state of the room for each tutor trial. Four sentences were added to the static picture of the form

The ball is on the chair. Only one was correct, and the other three had either incorrect fixed objects, movable objects, or prepositions. Foil types varied to prevent strategic guessing.

For both training conditions, correctness feedback was provided after every attempt. A participant could not advance to the next trial without providing the correct answer. Within-subjects, all participants completed a pre-test and an immediate post-test in both interfaces. Dependent variables were accuracy and latency of correct responses.

After each participant completed the two pre-tests, (one with the tutor interface and one with the multiple-choice interface), participants practiced for 40 minutes with either the tutor or multiple-choice interface. After training, participants completed a post-test in each format. The order of the two tests was counterbalanced and constant from pre- to post-test. In total, the experiment took about 90 minutes.

During pre- and post-tests, participants did not receive correctness feedback for their actions. In training, correctness feedback was always presented, and in the event of an incorrect response, participants had to provide the correct response before moving on to the next trial.

Results

We analyzed the test data for each test type separately, with a linear model including test time (pre-test / post-test), training condition (tutor / multiple choice), and the interaction plus an error term. Analyses were performed on two dependent variables: latency for correct responses and accuracy as proportion correct. For the accuracy analyses, we treat proportion correct as a binomial variable. Overall descriptive statistics are presented in Table 1.

-------------------------------TABLE 1 ABOUT HERE-------------------------------------

Pre-test performance. Before training, mean accuracy was .43 (*SD* = .50) for the more difficult tutor pre-test, and .65 (*SD* = .48) for the easier multiple-choice pre-test.   There was no difference in accuracy between training conditions at pre-test for either test type (for tutor pre-test, t (29) = 0.72, p = .48; for multiple-choice pre-test, t (29) = 0.06, p = .95).

A correct tutor trial took an average of 12.25s (*SD* = 10.35s), whereas a correct multiple-choice trial took 22.89s on average (*SD* = 16s). There was also no difference in response latency between training conditions at pre-test for either test type (for tutor pre-test, t (29) = 0.45, p = .66; for multiple-choice pre-test, t (29) = 0.16, p = .87)

Post-test performance. After 40 minutes of training, the overall mean proportion correct on the tutor post-test was .66 (*SD* = .48). For participants who trained with the tutor interface, average accuracy was .693 (*SD* = .46). For those who trained with the multiple-choice interface, average accuracy was .624 (*SD* = .49). A logistic mixed-effects model with a random intercept for each participant, with accuracy as a binomial dependent variable, showed improvement from pre-test to post-test across both training groups (β = 0.91, p < .001), and no overall main effect of training group (β = -0.44, p = .31). However, as predicted, tutor training led to greater average improvement from tutor pre-test to tutor post-test (β = 0.72, p = .001) than did multiple-choice training. Multiple-choice test data, with the same model, showed improvement across training groups from pre-test to post-test (β = 0.49, p = .005), and no overall main effect of training condition (β = -0.16, p = .74). Moreover, there was no difference in amount of learning for tutor training and multiple-choice training groups, showing that, for the multiple choice task, there was no advantage to training with the multiple-choice task itself (β = -0.13, p = .57). In separate mixed-effect models of latency of correct response (in seconds) for tutor and multiple-choice tests, there was a significant speed-up from pre-test to post-test for both test types (tutor test β = -3.76, $p < .001$; multiple-choice test β = -7.65, $p < .001$). In both cases, there was no effect of training group on the amount of speed-up (tutor test β = 0.79, p = .37; multiple-choice test β = 0.75, p = .58).

Discussion

Our first goal was to design and implement a computerized tutor capable of providing enactive practice in the use of prepositions for second language learners. The current study then provided initial evidence for the utility of this tutor. We predicted that an enactive practice environment would enable learners to make stronger representations of preposition words. The results of training showed greater learning with the tutor interface, when training was evaluated with the more difficult tutor test. This shows that practice with this enactive method was more effective for future performance than practice with multiple-choice selection.

One interpretation of the tutor training advantage on tutor interface post-tests is that it is simply the result of familiarity with the task. Familiarity surely leads to better performance. However, the lack of equivalent advantage for the multiple-choice training group on multiple-choice tests suggests that the benefit of training with the tutor interface is not entirely the result of task familiarization. Furthermore, it is reasonable to regard familiarization as a pedagogically relevant effect, because the task that learners completed in the tutor condition (following instructions moving objects) was closer to the usage of prepositions in real conversational contexts than was the task in the multiple-choice training and testing.

A limitation of the current study is the use of only one comparison group. Because of the complexity of the intervention, it is almost impossible to do a true minimal pairs comparison, with enactment as the only difference between groups. We controlled for the complexity of the verbal scene for the multiple-choice training group; however, the sentences in the multiple-choice trials were shorter, though because there were four sentences in the multiple choice condition compared to one in the tutor, each multiple choice trial took longer to complete. For future studies, it would be particularly useful to evaluate a test that no training group has seen to verify that there is an advantage to training with an enactive task that is reflected in unpracticed; that is, to confirm that enactive practice leads to either deeper levels of processing, stronger representations, or both. One possible test is sentence-picture verification (Clark & Chase, 1972), because the simplicity of the task would also allow a test of differences in speed-up and proceduralization after accuracy is already high. The addition of multiple training sessions and a delayed post-test would also be helpful to provide greater learning and examine robustness to forgetting.

In conclusion, this study describes the construction and initial test of a novel computerized tutor to designed to give students embodied practice using Spanish spatial prepositions in a virtual world context. The study illustrated that principles from a successful first-language reading comprehension intervention could be extended to apply to second language learning. The intervention out-performed a common instructional exercise of multiple-choice sentence-picture matching for the more complex test of preposition knowledge. This pattern is consistent with the previously demonstrated mechanism of enactive learning leading to deeper processing. This principle could have broad implications for language teaching and should be tested further.

References

Clark, H. H. & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,*472-517.

Coventry, K. R., Prat-Sala, M., & Richards, L. (2001). The interplay between geometry and function in the comprehension of "over", "under", "above" and "below". *Journal of Memory and Language, 44,* 376-398.

Coventry, K. R., & Garrod, S.C. (2004). *Saying, seeing and acting: The psychological semantics of spatial prepositions.* New York: Taylor & Francis.

Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2009). Improving early reading comprehension using embodied CAI. *Instructional Science, 39*, 27-39.

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology, 32*. 389-399.

Glenberg, A. M., Gutierrez, T., Levin, J.R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96,* 424-436.

Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379-401.

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*, 149-174.

MacWhinney, B. (in press). The logic of the Unified Model. In S. Gass and A. Mackey (Eds.), *Handbook of Second Language Acquisition*. New York: Routledge.

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary Educational Psychology, 32*, 537-550.

O'Malley, M. & Chamot, A. (1990). *Learning strategies in second language acquisition.* Cambridge: Cambridge University Press.

Regier, T., & Carlson, L. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, 130, 273-298.

Reiter, E. & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, 3(1).

		Pre-Tests		Post-Tests	
		Tutor	Multiple Choice	Tutor	Multiple Choice
Tutor Condition	Accuracy	.37 (.02)	.64 (.02)	.69 (.02)	.74 (.02)
	Latency	11.5 (0.53)	23.4 (0.75)	7.9 (0.22)	14.8 (0.46)
Multiple Choice Condition	Accuracy	.49 (.02)	.65 (.02)	.62 (.02)	.72 (.02)
	Latency	13.1 (0.54)	22.4 (0.74)	8.4 (0.23)	13.4 (0.58)

Table 1. Descriptive statistics (proportion correct and average response latency in seconds) for tutor and multiple choice groups across tutor and multiple choice tests before and after training.