The Role of Gestures in Spatial Explanations: Change in Spatial Perspective

Tatsuki Takenaga (takenaga@p.u-tokyo.ac.jp)
Graduate School of Education, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract

The goal of this paper is to investigate the role of gestures in verbal descriptions of a spatial layout. Participants listened to a text describing a fictional environment from the route or survey perspective, and then asked to explain the pairs of landmarks, some with route perspective (directions from one landmark to another) and other with survey perspective (to provide the locations of two landmarks). Analysis of gestures during the explanation in Experiment 1 revealed that speakers who learned from the route perspective produced more gestures when the explanation involved the survey perspective. The type of increased gestures was consistent with the perspective that the task demanded. These results support the Information Packaging Hypothesis. The fact that the effect of changing from survey perspective to route perspective was not found in Experiments 1 and 2 indicated an advantage of the survey text over the route text.

Keywords: gesture; perspective taking; spatial cognition; representation

Introduction

This paper investigates the underlying cognitive processes involved in relating spatial information from verbal input to two output modalities (i.e., speech and gesture). Recently, researchers have started exploring the possible cognitive function of co-speech gestures. The Information Packaging Hypothesis (Kita, 2000) has been proposed to explain how gestures and speech are produced. According to this hypothesis, gestures help speakers to organize knowledge that is spatio-motoric in nature and convert it into a verbally expressible form (e.g., propositional representation). If gestures are in themselves beneficial, they increase with the difficulty of the problem.

Some studies have provided supportive evidence for the Information Packaging Hypothesis, by varying the difficulty of conceptualization. For example, Hostetter, Alibali, & Kita (in press) found that speakers produced more gestures when the task required complex conceptualization. Alibali, Kita, Young (2000) found the same result in children. Moreover, many studies showed that the rate of gestures was related to the cognitive load (Gaber & Goldin-Meadow, 2002; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). These results were also consistent with the Information Packaging Hypothesis.

However, the types of gestures reflecting cognitive load are unclear. Although Emmorey, Tversky, & Taylor (2000) showed that the type of gesture was compatible with the type of speaker's description, their task did not manipulate the cognitive load. In this study, the rate and type of gestures were analyzed using a spatial explanation task that involved a different perspective.

According to studies on large-scale spatial representations, two perspectives of spatial representation have been proposed—route perspective and survey perspective. The route perspective is viewpoint within an environment. On the other hand, the survey perspective is outside the environment. Some studies have revealed that there exists a cost for switching perspectives from encoding to test (e.g., Shelton & McNamara, 2004; Taylor & Tversky, 1992). They use text descriptions followed by text-based questions regarding the spatial layout. Such texts rely on spatial terms as the medium for route and survey perspectives (e.g., right–left vs. north–south). Comparison of the performance for a "same perspective" (e.g., route study–route test) and the performance for a "different perspective" (e.g., route study–survey test) revealed the cost for perspective switching. In this study, by modifying the paradigm, participants also learned a fictional environment via the text of either the route perspective or survey perspective, and then explained routes and the location of landmarks. If there is a cost of switching one perspective to another, the rate of gestures for different perspective (e.g., route study–location explanation) should be higher than the rate of gestures for a same perspective (e.g., survey study–location explanation).

In terms of the type of gestures reflecting cognitive load, there are two theoretical possibilities. The lexical retrieval hypothesis proposed that gestures are generated from the semantics of lexical items (Morrel-Samuels & Krauss, 1992). Morsella & Krauss (2004, 2005) showed that the concreteness and spatiality of words is associated with gesture production. The suggestion of these studies assumes that gestures have the iconicity of the spatio-motoric nature of semantic representation, and are generated from pre-existing representation. Moreover, Franklin (2007) showed that speakers produced gestures reflecting factual representation when they asked to deceive a listener by telling a story. If gestures are generated on such a pre-existing mental model already constructed by the speakers, then the type of gesture should always be compatible with the type of spatial representation that the speakers form. According to these pre-existing views of representation, in location explanation, participants who learned from the route perspective should produce *Enactment mode gestures* (more route-oriented gestures), that correspond to the learned perspective, than those who learned the survey perspective. In route explanation, on the other hand,

participants that learned from the survey perspective should produce more survey-like gestures (referred to as *Objectification mode gestures*), that correspond to the task, than those who learned from route perspective.

The Information Packaging Hypothesis, on the other hand, proposed that gestures originate from an impromptu representation that results from a "thinking for speaking process." If so, the type of gestures should align with a perspective that has an advantage for performing the task. According to the Information Packaging Hypothesis, in location explanation, participants that learned from a route perspective should produce more objectification mode gestures that correspond to the task, than participants who learned the survey perspective. In the route explanation, on the other hand, participants that learned from a survey perspective should produce more enactment mode gestures that correspond to the task, than the participants that learned from route perspective.

Experiment 1

Conceptualization in a large-scale spatial representation is considered to involve a switching of perspectives. If gestures do indeed help in conceptualization, speakers should produce more gestures that are consistent with the perspective that the task demands. Consequently, verbal descriptions of a fictional environment were used as stimuli in order to gain control over the amount of route and survey learning. To confirm that participants had in fact built up corresponding mental representation, a map-drawing task was used.

Method

Participants Thirty-two native Japanese speakers (20 males, 12 females) were recruited from an undergraduate educational psychology class at the University of Tokyo. Participants were randomly assigned to a survey text group or route text group, with the constraint that each group contain equal numbers of male and female participants.

Texts Two types of texts, modified from a version of Shelton & McNamara (2004), were prepared. Both involved a fictional environment, one based on a route perspective, and the other, on a survey perspective (see Figure 1). The survey text first introduced the four major quadrants of the environment, and then, explained the individual landmarks. The route text, on the other hand, immediately introduced the first landmark and revealed information on the overall layout of the environment in a stepwise manner. Furthermore, the survey text employed *canonical spatial terms* such as *in the southern part of*, while the route text used *relative spatial terms,* such as *on your left*.

Explanation tasks Participants were required to explain 15 pairs of landmarks (i.e., 9 route explanations and 6 location explanations). In the route explanation task, which demanded the route perspective, they were asked to give directions from one landmark to another. In the location explanation task, which demanded the survey perspective, they were asked to explain the locations of two separate landmarks.

Map Drawing After completing the explanation tasks, participants were given a blank sheet of paper and asked to sketch the environment they had learned. No particular orientation was required for the map drawing. The order of drawing landmarks and the choice of drawing orientation was analyzed.

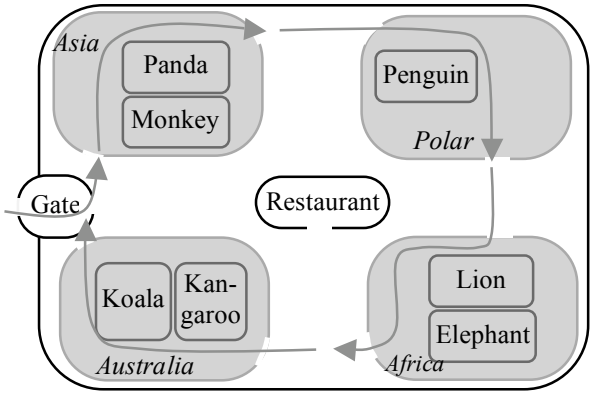

Figure 1: The map of fictional environment[1]

Procedure

Half of the 32 participants were asked to listen to the route text, while the remaining half listened to the survey text. The participants listened to the text up to four times. They were asked to memorize the text as much as possible. After the encoding phase, they completed two types of explanation tasks (i.e., route explanation and location explanation). The order of the two tasks was counterbalanced over the participants. Following the explanation tasks, the unaware participants were asked to draw a map of the environment. With the permission of participants, the entire experiment was recorded for analysis.

Gesture coding Each gesture in the explanation tasks was classified into one of two modes (i.e., *Objectification Mode* and *Enactment Mode*). Objectification mode gestures were characterized using a 2D plane as if the speakers were outside the environment, and were responding from a survey perspective. In particular, speakers produced gestures as though drawing a map on a horizontal plane or on a vertical plane.

Enactment mode gestures appeared as if the speakers were within the environment and indicated directions from their own perspective (i.e., route perspective). Enactment mode gestures generally moved away from the speaker and used the entire 3D space (rather than a 2D plane).

A single coder initially coded all of the data. For the reliability check, two coders who were blind to the

[1] Note that participants were not presented with the map.

experimental hypotheses independently assessed another 12.5% of the data (the explanation task from each 4 participants). Agreement between coders was 89% for classifying the mode of gestures (into objectification mode or enactment mode).

Results

Gestures during explanation tasks The proportion of responses accompanying one or more gestures was subjected to an analysis of variance, with terms for text (route vs. survey) as a between-subjects factor, and mode of gesture (objectification vs. enactment), as a within-subjects factor.

Gestures in location explanation If gestures are related to changing perspective, the route text group should produce more gestures than the survey text group. The main effects of the mode of gesture (F (1,30) = 38.60, $p < .001$) and text (F (1,30) = 9.28, $p < .01$) were significant. According to the Information Packaging Hypothesis, the route text group, should produce more objectification mode gestures, than the survey text group, with regard to location explanation. For the objectification mode gesture, the simple main effect of text was significant (F (1,30) = 11.18, $p < .05$, see the left part of Figure 2).

Gestures in route explanation The main effect of mode of gesture was again significant (F (1,30) = 21.51, $p < .001$). However, the main effect of text was not significant (p = .18). According to the Information Packaging Hypothesis, the survey text group, should produce more enactment mode gestures, than the route text group, with regard to route explanation. For the enactment mode gesture, the simple main effect of text was significant (F (1,30) = 5.13, $p < .05$, see the right part of Figure 2). However, the effect of text was opposite of that expected: the survey text group produced fewer enactment mode gestures.

Figure 2: Proportion of gesture – accompanied responses in explanation tasks

Use of spatial terms during explanation tasks In order to control the language used between the groups, the total number of spatial terms and the proportion of spatial terms in explanation tasks were analyzed. Although the survey text group produced slightly more spatial terms than route text group, the group effect was not significant (survey text (M = 53.4 vs. route text M = 62.2, F (1,30) = 0.90, $p > .1$). Thus, there was no evidence for the difference of the rate of gestures due to the amount of spatial terms.

The spatial terms used in the explanation tasks were classified as canonical spatial terms (e.g., *east* and *west*) or relative spatial terms (e.g., *in front of you*).

For the location explanation task, the survey text group used more canonical spatial terms (M = 97%) than the route text group (M = 79%, F (1, 30) = 4.79, $p < .05$, see Figure 3). According to the lexical retrieval hypothesis, the survey text group that used more canonical spatial terms should produce more objectification mode gestures. However, the result obtained was the opposite of that expected. The survey text group that used more canonical spatial terms produced fewer objectification mode gestures.

For the route explanation task, the route text group used more relational spatial terms (M = 84%) than the survey text group (M = 12%, F (1, 30) = 63.61, $p < .001$, see Figure 3). The results of spatial terms suggested that participants constructed corresponding spatial mental model.

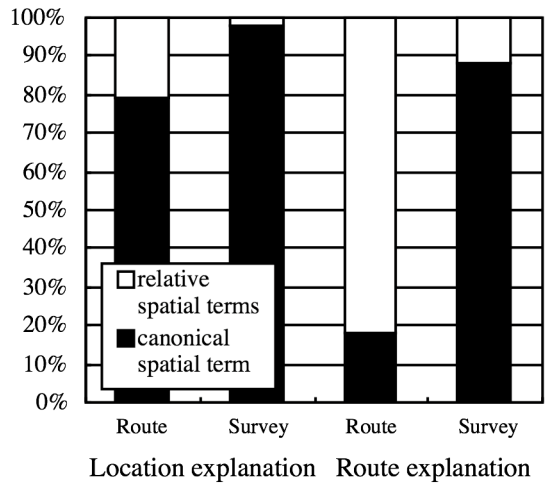

Figure 3: Proportion of spatial terms used in explanation tasks

Further analysis of gesture and speech in route explanation The effect of switching perspectives was not detected in route explanation, because most of the spatial terms used in the survey text group were canonical spatial terms (see Figure 3). To provide further evidence for the Information Packaging Hypothesis, gesture production by participants who used only spatial terms *congruent* with the text was compared with that of participants who used *both* spatial terms (relative and canonical spatial terms) in route explanation. Participants who used both spatial terms could

partly change from route perspective to survey perspective in route explanation. Participants who used only congruent spatial terms, on the other hand, did not change their perspective. According to the Information Packaging Hypothesis, participants who use both spatial terms should produce more enactment mode gestures than those who use only congruent spatial terms. The data from route text group supports this hypothesis (F (1,30) = 4.46, $p < .05$, see Table 1). In terms of objectification mode gestures, there was no difference between two groups ($p > .1$).

Table 1: Proportion of enactment mode of gesture – accompanied responses in the route explanation task (route text group)

Enactment mode of gestures	
Both spatial terms	Congruent spatial terms
0.46	0.12
(0.44)	(0.24)

Map drawing To confirm that participants constructed corresponding spatial representation, the order of drawing landmarks and drawing orientation were analyzed. In route representation, information about landmarks was organized in a stepwise manner. In survey representation, on the other hand, information about landmarks was organized in a hierarchic manner. Thus, the temporal order of landmarks was more important in route representation than in survey representation. If the text affected spatial representation that participants formed, the order of drawing landmarks should be affected by the text in the route text group rather than survey test group. Rank correlation between the order of drawing landmarks and the order of presented landmarks in the text was calculated. The mean correlation coefficient was 0.86 for the route text group and 0.29 for the survey text group. There was significant difference between two groups (t (24) = 7.38, $p < .001$).

The drawing orientation was also analyzed. All participants in survey text group drew the map from north-as-up orientation. Eleven out of 16 participants in the route text group drew the map from the initial orientation of the text description. Thus, they drew the map from north-as-left orientation. The other participants in route text group drew the map from north-as-up orientation. The drawing orientation was a significant difference between two groups ($p < .001$).

These results of map drawing indicated that participants constructed the spatial mental model that was consistent with the perspective of the text.

Discussion

When the route text group used canonical spatial terms, they produced more gestures, than the survey text group or those who used only relative spatial terms. Moreover, in terms of the mode of gestures, gestures produced were consistent with what the task demanded rather than the spatial terms used.

These results support the Information Packaging Hypothesis in that the gestures demanded by the task assist in conceptualization. The gestures help participants switch perspectives as they use spatial representation congruent with the demands of a specific task. However, in the current experiment, the survey text group did not produce more gestures in route explanation than the route text group. One possible reason is that the survey text led to a spatial representation with more fine-grained localization than the route text. Another possible reason is that the survey text group hardly used relative spatial terms. Further experiment was required to explore these possibilities.

Experiment 2

In Experiment 1, participants freely chose their own spatial terms. It is possible that the survey text group might not have changed their perspective in the route explanation task, because most of the spatial terms used in the survey text group were canonical spatial terms (Figure 3). To test the significance of canonical spatial terms, the instruction in the spatial explanation task was changed. In Experiment 1, participants could freely choose canonical spatial terms. In Experiment 2, participants were asked to use relative spatial terms.

Method

Participants Sixteen native Japanese speakers (8 males, 8 females) were recruited from an undergraduate educational psychology class at the University of Tokyo. All participants were assigned to the survey text group.

Text Survey text was the same as in Experiment 1.

Explanation tasks The instruction of how participants explained the route direction and the location was modified. Participants were required to explain 15 pairs of landmarks (i.e., 9 route explanations and 6 location explanations). In the route explanation task, which demanded the route perspective, they were asked to give directions from one landmark to another, using *relative spatial terms* as much as possible. In the location explanation task, which demanded the survey perspective, they were asked to explain the locations of two landmarks, using *canonical spatial terms* as much as possible.

Procedure This remained the same as with the survey text group in Experiment 1, except for a modification of explanation tasks.

Results

Spatial terms in explanation tasks To check the manipulation of instruction, the proportion of spatial terms in explanation tasks was analyzed (see Figure 4). In the location explanation task, the survey text group used canonical spatial terms (99%). In the route explanation task,

the survey text group in Experiment 2, used more relative spatial terms (M = 57%) than in Experiment 1 (M = 12%, F (1,15) = 18.92, p < .001). This result showed that the modified instruction successfully led to the increase of relative spatial terms in the route explanation.

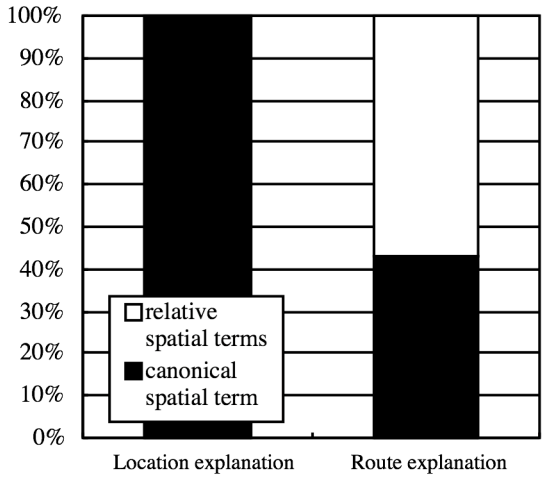

Figure 4: Proportion of spatial terms in survey text group in Experiment 2

Gestures during explanation tasks Coding of gestures in spatial explanation tasks was the same as in Experiment 1.

Gestures in location explanation The survey text group produced more objectification mode of gestures than enactment mode of gestures in location explanation (F (1,15) = 18.92, p < .001, see the left part of Table 2).

Gestures in route explanation The survey text group produced more objectification mode of gestures than enactment mode of gestures in route explanation (F (1,15) = 25.72, p < .001, see the right part of Table 2).

Table 2: Proportion of gesture – accompanied responses in explanation tasks (survey text group)

Location explanation		Route explanation	
Objectification	Enactment	Objectification	Enactment
0.43	0	0.53	0.08
(0.42)	(0)	(0.42)	(0.15)

Parenthetical values are between-subjects standard deviations

Comparison with Experiment 1 and Experiment 2
To test the significance of the using canonical spatial terms, the survey text groups in Experiment 1 and in Experiment 2 were compared. If the use of canonical spatial terms was important, the modified instruction in Experiment 2 should lead to an increase in the rate of gestures. When compared, the survey text group in Experiment 1 and in Experiment 2, showed no difference in the proportion of gestures in the route explanation task (F (1,30) = 0.59, n.s.). Moreover, the route text group in Experiment 1 produced more enactment gestures than the survey text group in Experiment 2 (F (1,30) = 6.55, p < .05). Experiment 2 replicated the data of the previous experiment's survey text group.

Discussion

The purpose of Experiment 2 was to test the significance of canonical spatial terms. However, the result of the survey text group in Experiments 1 and 2 was quite similar. This result implied that the use of spatial terms was not important, at least in the changing from survey perspective to route perspective. The cost of perspective switching was not detected in Experiment 1, because of the superiority of survey representation to route representation. Spatial representation from survey perspective may be more fine-grained and general than route perspective.

General Discussion

Using spatial explanation tasks, speech protocols were collected in which participants could freely choose whether to produce gestures. The goal of this paper is to analyze the rate and type of gestures, yielded by the spatial explanation task that involved a change in perspective. When the route text group used canonical spatial terms, they produced more gestures than either the survey text group or those who used only relative spatial terms. Cognitive load of switching from route perspective to survey perspective led to an increase of gestures in spatial explanation tasks. In terms of the types of gesture, the route text group produced more gestures consistent with perspectives that have an advantage in performance of a task. These results supported the Information Packaging Hypothesis. The gestures help speakers switch perspectives as they reorganize spatial mental model congruent with the demands of a specific task.

In terms of changing from survey perspective to route perspective, there was no significant increase of gestures. Comparison of the survey text group in Experiment 1 and in Experiment 2, implied that the cognitive load of switching survey perspective to route perspective, was slightly smaller. According to more recent studies on large-scale spatial representation, survey description was found to be superior to route description (Noordzij & Postma, 2005; Noordzij, Zuidhoek, & Postma, 2006). In distance comparison task, participants who learned from the route perspective made more errors than those who learned from the survey perspective (Noordzij, Zuidhoek, & Postma, 2006).

The result of map drawing showed that participants constructed the spatial representation that was consistent with the perspective of the texts. This result might have reflected the static nature of the spatial representation. Gestures generated from an impromptu, dynamic representation that might be separated, form a situation model. A situation model, that speakers produced gestures, is related to the proposal of the role of perceptually represented structure in language comprehension. Barsalou

proposed a Perceptual Symbol Systems theory, that language comprehension can be explained as the language-induced mental simulation of the described actions (Richardson, Spivey, Barsalou, & McRae, 2003, see also Barsalou, 1999).

Shelton & McNamara (2004) showed the interesting finding that there was orientation dependence in route and survey leaning. Through scene recognition for route and survey images, when the test view matched the participants' learned orientation, the same perspective recognition performance was facilitated. Kita (2003) investigated gestures during the route description in the real environment. Although his task did not manipulate the perspective, he proposed that the choice between the concepts right and left during speech was facilitated by the alignment of gestures and the speaker's torso orientation with the environment. The expansion of these studies would consist of examining whether there are systematic relationships between orientation dependence in the same perspective and gesture productions.

Acknowledgments

A part of this research was presented previously at the 5th International Conference of the Cognitive Science (Takenaga, 2006).

References

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15,* 593–613.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

Emmorey, K., Tversky, B., & Taylor. H. A. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, *2,* 157-180.

Franklin, A. (in press) Blending in Deception: Tracing Output Back to its Source. in Duncan, S. D., Cassell, J. &. Levy, E. T. (Eds.) *Gesture and the Dynamic Dimension of Language: Essays in Honor of David Mcneill*

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action*, pp.162–185. Cambridge, UK: Cambridge University Press.

Kita, S. (2003). Interplay of gaze, hand, torso orientation, and language in pointing. In Kita, S. (Ed.) *Pointing: Where language, culture, and cognition meet*. pp. 307-328. Mahwah, NJ: Lawrence Erlbaum.

Hostetter, A. B., Alibali, M. W., Kita, S. (in press). I see it in my hands' eye: Representational gestures reflect conceptual demands. . *Language and Cognitive Processes.*

Garber, P. & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science*, *26*, 817-831.

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*, Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12,* 516-522.

Morrel-Samuels, R., Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning Memory & Cognition, 18*, 615-622.

Morsella, E, & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, *117*, 411-424.

Morsella, E, & Krauss, R. M. (2005). Muscular Activity in the Arm during Lexical Retrieval: Implications for Gesture-Speech Theories. *Journal of Psycholinguistic Research, 34*, 415-427.

Noordzij, M. L., & Postma, A. (2005). Categorical and metric distance information in mental representations derived from route and survey descriptions. *Psychological Research, 69*, 221-232.

Noordzij, M. L., Zuidhoek, S., & Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition, 100*, 321-342.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767-780.

Shelton, A. L., & McNamara, T. P. (2004). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning Memory, & Cognition, 30*, 158–170.

Takenaga, T. (2006). The role of gestures in spatial representations derived from route and survey texts, *Proceedings of The 5th International Conference of the Cognitive Science*, 203-204.

Taylor, H. A., Tversky, B. (1992). Spatial Mental Models Derived from Survey and Route Descriptions, *Journal of Memory Language*, *31*, 261-292.

Taylor, H. A., Tversky, B. (1996). Perspective in Spatial Descriptions, *Journal of Memory Language*, *35*, 371-391.

Wesp, R., Hesse. J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery *American Journal of Psychology. 114*, 591-600.