Title
Spatial Preposition Use in Indoor Scene Descriptions

Journal
International Conference on GIScience Short Paper Proceedings, 1(1)

Authors
Doore, Stacy
Beard, Kate
Giudice, Nicholas

Publication Date
2016-01-01

DOI
10.21433/B3115h00g1xs

Peer reviewed

Spatial Preposition Use in Indoor Scene Descriptions

S. A. Doore, M.K. Beard, N.A. Giudice

[1]Spatial Informatics, School of Computing and Information Science, University of Maine
Email: {stacy.doore@maine.edu, beard@spatial,maine.edu, nicholas.giudice@maine.edu}

Abstract

In order to provide accurate automated scene description and navigation directions for indoor space, human beings need intelligent systems to provide an effective cognitive model. Information provided by the structure and use of spatial prepositions is critical to the development of accurate and effective cognitive models. Unfortunately, the use and choice of spatial prepositions in natural language is extremely varied, presenting difficulties for natural language systems attempting to provide descriptions of indoor scenes and wayfinding directions. The goal of the present study is to better understand how humans use spatial prepositions to communicate spatial relationships within virtual environment (VE) indoor scenes. A series of experiments investigates spatial preposition use and the influence scale, topology, orientation and distance within indoor scene descriptions and preliminary results are reported.

1. Introduction

Humans perceive and represent information differently for indoor spaces than they do for outdoor spaces (Guidice, Walton and Worboys 2010). Indoor spaces usually lack established distance and direction metrics, global landmarks, explicit route-networks, and cover a range of spatial scales from small rooms to large airports (Winter 2012). Our interest is in creating an indoor spatial description system to assist navigation in indoor environments. This research investigates natural language (NL) structures for describing objects and structural features within *vista scale* (Montello 1993) virtual environment (VE) indoor scene descriptions. We examine how spatial preposition choice may vary in different contextual settings and which spatial prepositions might yield more effective spatial representations with an increased ability to support spatial behaviors (e.g., object location and navigation).

This paper describes early work investigating how spatial factors such as room size, scene elements, and object/structure relationships impact spatial preposition use and semantics for indoor scene descriptions. Previous research has found these same variables of topology, scale, orientation and distance impact spatial preposition use in geographic and table top spatial settings. The current study focuses specifically on vista scale VE scenes (i.e., 10'by 12' and 20' by 30') that one might encounter in common built environments (e.g., home, business, or school size rooms).

2. Motivation

Consider two spatial expressions describing the same indoor scene:

Speaker 1: *"The bookcase is against the wall, beyond the table, just to your left."*
Speaker 2: *"The bookcase is to the left of the window, directly in front of you."*

In both cases, the speakers use the bookcase as the focus of the spatial description, however, the statements differ in how the bookcase is spatially situated in the scene. The first speaker uses the wall as the primary spatial landmark, whereas, the second speaker uses the window.

The spatial prepositions used by the speakers also differ. Although both expressions are semantically correct, the differences in the linguistic structure of the sentences will yield different cognitive maps for someone trying to mentally reconstruct the scene solely based on each individual description. For most people, scenes are perceived visually. However, in the absence of visual support, communicating a spatial description is an error prone process with a high probability of information uncertainty (e.g., vagueness, ambiguity, inaccuracy, etc.).

Humans construct cognitive maps, or allocentric, global representations of space, that are specialized to individual needs, sensory capacities, and tasks that support spatial inference as well as route planning and way-finding (Downs and Stea 1973). While many physical maps and verbal scene/route descriptions leave out or distort spatial detail, this tends to be the same information that is also omitted/distorted in cognitive maps, such as metric information about distance and direction (Tversky 2001). The goal of this study is to better define the types of descriptive spatial detail necessary to fill-in the perceptual cues supporting accurate cognitive map construction of indoor scenes.

3. Methods

A series of three experiments were conducted to investigate the contextual factors of spatial preposition use for a meets/touches relation within a highly controlled virtual environment (VE). The VE scenes allowed for a more controlled environment for isolating the specific scene elements of interest (topology, orientation, distance). An earlier study on this topic found no significant difference between real world and virtual world scenes descriptions (Kesavan and Giudice 2012). The experiments address the following questions in order to better define formal rules for a future indoor spatial description generation system:

- What are minimally descriptive but sufficiently effective NL descriptions for specifying spatial relations between entities within indoor environments when considering different types of context dependencies?
- What are critical factors that influence spatial prepositions choice/preference within different scale indoor environments (i.e., room-size, object-type, orientation, distance)?

One hypothesis of this study is that underspecified spatial prepositions (e.g., on, in, at, by) can be used effectively as spatial information communication short cuts in verbal indoor scene descriptions because for sighted users, unlike their blind or low BLV peers, a high level of verbal spatial detail is unnecessary to create effective cognitive maps. These simple spatial prepositions are the first to emerge during language acquisition between the ages of 1-5 and are overgeneralized to represent a variety of spatial relationships at this early language development stage (Clark 1973). We believe that simple spatial prepositions can be effectively overgeneralized in indoor scene descriptions in much the same manner as other types of spatial details are reduced in outdoor geographic settings (e.g., GPS route directions) in order to reduce the cognitive load of the description recipient in constructing a minimally descriptive but sufficiently effective cognitive map.

In the first experiment, participants were asked to respond to a series of prompts about objects and room structures in VE indoor scenes. The prompts required participants to provide missing spatial prepositions in both text and verbal response formats to describe the spatial configuration of the specified objects (e.g., desk, chair, bookcase) and room structures.(e.g., wall, door, window). Half of the participants were asked to create these short spatial descriptions for a hypothetical person who could see the VE scene themselves and the other half of participants were asked to create the short descriptions for a hypothetical BLV user. This methodological approach has been used in a similar study investigating NL descriptions of spatial relations in outdoor geographic space (Klippel, Weaver and Robinson 2011).

In the second experiment, participants classified sets of five images of similar indoor scenes into 3 unlabeled groups or 4 pre-determined categories based on their evaluation of the most appropriate spatial preposition to represent the meets/touches or disjoint spatial relations in the VE scenes. This is a similar approach used in previous studies investigating NL representation of scene elements in geographic space (Klippel, Weaver and Robinson 2011). The final experiment asked participants to look at a VE scenes and then evaluate a spatial prepositions used for a meets/touches relation based on three scales: similarity, clarity, and preference (Schwering 2007).

4. Results

Preliminary results suggest a strong preference for the use of *on* for the meets/touches in both verbal and text response sets. Additional room structures were identified in the structured prompt responses, such as "corner" and the "middle" in the descriptions of object-structure relations within the VE scenes. These room structures worked as *containment* structures for objects when a clear *meets/touches* relation was not possible because of a clear disjoint relation with the object and room structure in question. We believe that these types of elementary spatial prepositions are more frequently used to describe object and room structure relations because, in most cases, more precise spatial information contained in more descriptive spatial prepositions (e.g., along, parallel, perpendicular, etc.) is not necessary for those who can access their vision as a primary sensory modality. A full analysis of the results is currently underway and will employ linguistic, machine learning, and cluster analysis techniques.

5. Conclusions

The results of this research will provide new information about the level of semantic precision necessary for describing indoor space in a manner robust enough to support spatial learning, cognitive map development, and spatial behaviors as well as real-time semantic scene searching. Practical applications of this research include more effective NL spatial language formalisms that can be used in the areas of assistive technologies, emergency response, location-based services and homeland security, which all require the accurate communication of fine scale indoor spatial knowledge.

Acknowledgements

This research was supported by NSF grant CDI-1028895.

References

Clark H H, 1973, Space, time, semantics and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press. Coleman.

Downs R and Stea,D, 1973, *Image and the Environment*. Aldine, Chicago.

Guidice, N A, Walton, L A, and Worboys, M A, 2010, The informatics of indoor and outdoor space , (eds.) *ISA 2010.Proceedings of 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, ACM, New York, 47-53.

Kesavan S, Giudice N A, 2012, Indoor scene knowledge acquisition using a natural language interface. In Graf C Giudice, N Schmid F (eds) *SKALID 2012. Spatial Knowledge Acquisition with Limited Information Displays*, Kloster Seeon, Germany,1-6.

Klippel A, Weaver C, and Robinson, A C, 2011, Analyzing cognitive conceptualizations using interactive visual environments. *Cartography and Geographic Information Science*, 38,1, 52–68.
Montello D, 1993, Scale and multiple psychologies of space. In Frank A and Campari I (eds) *Spatial Information Theory: A Theoretical Basis for GIS*. *LNCS* 716, Springer-Verlag, New York, 312-321.
Schwering, A, 2007, Evaluation of a semantic similarity measure for natural language spatial relations. In Winter S Kuipers, B Duckham M and Kulik L (eds), *Spatial Information Theory. 9th International Conference, COSIT 2007*, Melbourne, Australia, Berlin: Springer, 116-132.
Tversky B, 2001, Spatial schemas in depictions. In Gattis M (ed), *Spatial Schemas and Abstract Thought.* MIT Press: Cambridge, 79-111.
Winter S, 2012, Indoor spatial information. *International Journal of 3-d Information Modeling*. 1,1. 25-42.