

The Impact of Dimensionality on Natural Language Route Directions in Unconstrained Dialogue

Vivien Mast University of Bremen Bremen, Germany viv@tzi.de	Jan Smeddinck University of Bremen Bremen, Germany smeddinck@tzi.de	Anna Strotseva University of Bremen Bremen, Germany anna.strotseva@uni-bremen.de	Thora Tenbrink University of Bremen Bremen, Germany tenbrink@uni-bremen.de
---	---	--	--

Abstract

In this paper we examine the influence of dimensionality on natural language route directions in dialogue. Specifically, we show that giving route instructions in a quasi-3d environment leads to experiential descriptive accounts, as manifested by a higher proportion of location descriptions, lack of chunking, use of 1st person singular personal pronouns, and more frequent use of temporal and spatial deictic terms. 2d scenarios lead to informative instructions, as manifested by a frequent use of motion expressions, chunking of route elements, and use of mainly 2nd person singular personal pronouns.

1 Introduction

In order to build artificial agents that are competent in creating and understanding natural language route directions in situated discourse, it is necessary to explore how situatedness affects the communication of humans about routes. The current study aims at exploring in which ways dimensionality influences the choice of communicative strategies for route directions in discourse.

Previous research about route directions mostly deals with monologues or pretend dialogue (e.g. Rehr et al., 2009; Klippel et al., 2003), and concerns two-dimensional stimuli, such as map-based tasks (Klippel et al., 2003; Goschler et al., 2008).

The study presented here examines pairs of participants collaborating on a route instruction task in a naturalistic discourse setting under two conditions: In the 2d condition, the instructor was shown a two-dimensional map with the route drawn into it. In the 3d condition

however, the instructor navigated along a pre-set route in Google Maps Street View.

2 Route Instruction Strategies

Route directions consist of *procedures* and *descriptions* that combine to a step-by-step prescription of the actions that are necessary for executing the given course (Michon and Denis, 2001; Longacre, 1983). Since spatial linguistic expressions reflect the mental model already existing on the part of the instructor, the dimensions in which route instructors experience an environment (2d or 3d) may have a systematic impact on the discourse strategies they use. In the following we analyze a range of spatial descriptions, focusing on aspects known to be crucial for spatial interaction, such as descriptions of locations and motion, the use of perspective expressions, chunking of route elements, and personal and spatiotemporal deixis.

2.1 Static and Dynamic Descriptions

Since route directions deal with a static environment in which a movement takes place, they usually include a high proportion of dynamic descriptions of actions (*procedures* in Michon and Denis' (2001) terms), and additional static information about the surroundings (*descriptions*). In our analysis, we distinguished speakers' utterances as *motion descriptions* if they described or requested the literal *motion* of an entity. In contrast, an utterance was marked as *location* if it described a *static spatial relation*, for example the position of the speaker or an object at a certain point in time.

2.2 Perspective Use

When describing routes, speakers either use the *route perspective*, describing route elements or motions from the point of view of a

person traveling along the route, or the *survey perspective*, where the description is based on cardinal directions, or directions as they are defined by the map as a whole (Taylor and Tversky, 1996). Previous research has indicated that perspective choice can be influenced by the specific situation, and by the coordination between speakers in natural discourse (Pickering and Garrod, 2004; Watson et al., 2006). In the present study, we test the hypothesis that navigating a route in a 3d perspective makes it more difficult for the instructor to use the survey perspective, leading to a preference for the route perspective. Further we assume that the follower will adapt to the instructor's perspective choice in terms of language use.

2.3 Chunking of Route Segments

In a study examining online route descriptions to an imaginary follower based on a two-dimensional map, Klippel et al. (2003) found that participants tended to chunk decision points without directional change together. For example, a speaker could say “turn right at the second intersection” instead of “Go straight on, and then turn right”. This occurred even when the route was shown as a moving dot on the map. In our study, we address the question whether this also holds for instructors with a three-dimensional view. We expected a frequent usage of chunking in the 2d condition, in which the participants have access to comprehensive structural information, as opposed to a higher degree of separate references in the 3d condition, in which participants experience the environment incrementally.

3 Experiment

22 students (average age 25, 14 male and 8 female) volunteered to participate in the experiment. They formed 11 pairs that each completed one test run and three permuted critical trials. Instructor and follower were placed in different rooms and interacted via telephone software.

The four predetermined routes were identical for all participants, and they differed mildly in complexity, ranging from 9 to 14 decision points. All routes were located in San Francisco and were specifically designed such that, at most decision points, descriptions would be unambiguous with respect to perspective use.

In the 2d condition (5 pairs), instructors were given a map that showed mostly street names and major landmarks such as parks, schools, restaurants, etc., as they appear in the standard Google Maps map view. The route consisted of a marked starting and end point, and was signaled by a thick blue line with arrows indicating the direction. In the 3d condition (6 pairs), instructors interacted directly with Google Street View which had a photographic quasi-3d view and allowed them to observe the surroundings as if navigating on the roads, seeing a vast amount of details of the environment. Street names were clearly readable as an overlay on top of the photographic imagery. The route was indicated by fat blue arrows that the instructors could click on, in order to move in the given direction.

In both conditions, the followers were asked to draw the route on a map that only contained the starting point. The task instruction was the same for both conditions, priming for procedural discourse yet ambiguous with respect to perspective use: “Now you have to tell your partner where you are going. Please do this by giving instructions via the microphone.” (translated from German). In the 3d condition instructors were informed that the follower had a different view of the same surroundings.

Taken together this setup differs from previous studies in that it features unconstrained spoken dialogue and is set in a realistic use-case with a three-dimensional setting.

4 Results

The participants in the 3d condition took significantly longer ($M = 125.61$ utterances per trial) to complete a task than the participants in the 2d perspective ($M = 46.40$ utterances per trial, $t(9) = 4.781$, $p = 0.001$).

Figure 1 shows typical examples of the instructors' language in the two conditions. In the 2d condition, instructors as well as followers used survey perspective, as in line 2.2 in Figure 1, significantly more frequently than in the 3d condition (see Table 1). A Chi-square test showed the following results for the instructors: $\chi^2(1) = 200.14$, $p < 0.0001$ and $\chi^2(1) = 91.25$, $p < 0.0001$ for the followers¹. It is notable that the followers in the 2d condition showed a preference for survey perspective ($\chi^2(1) = 15.38$; $p < 0.0001$), while in the 3d

¹ Mixed, conflated and unclear expressions were excluded from the analysis.

condition they clearly favored route perspective, which was the perspective of the instructor.

3d condition:

1.1 Yes... erm ...
now ... there is a crossing again

1.2 Moraga Street

1.3 to the left

1.4 into Moraga Street [...]

1.5 then there is a crossing again

1.6 the twelfth

1.7 straight on over there

1.8 So Moraga further

2d condition:

2.1 And then we go down that one up to Moraga Street

2.2 And there we also go right into Moraga Street

2.3 We go through that one up to Eleventh Avenue

Figure 1. Typical examples of instructors' language in the two conditions.

The instructors in the 3d condition used a significantly higher proportion of location descriptions than the instructors in the 2d condition ($t(6.5) = 4.500, p = 0.003$). As Table 2 shows, the instructors in the 2d condition relied mainly on motion descriptions (see Figure 1, location descriptions in lines 1.1 and 1.5 as well as motion descriptions in lines 2.1-2.3).

Perspective expressions	3D		2D	
	Instructor	Follower	Instructor	Follower
Route	98.93% (370)	93.33% (112)	50.88% (87)	21.57% (11)
Survey	1.07% (4)	6.67% (8)	49.12% (84)	78.43% (40)
Totals	374	120	171	51

Table 1. Use of perspective expressions in 2d and 3d conditions (absolute values in parentheses).

Chunking of route elements did not occur at all in the 3d condition. In the 2d condition there were 29 intersections that were skipped through chunking, as shown in line 2.3 in Figure 1. This amounts to a mean of 1.9 chunked intersections per route.

Instructors in the 2d condition strongly preferred 2nd person singular pronouns, whereas

instructors in the 3d condition showed a preference - though not as strong - for 1st person singular (see Figure 2). Instructors in the 3d condition also used the German formal pronoun *es* 'it' more frequently than those in the 2d condition. This preference is usually displayed in utterances noting the presence of landmarks in the surroundings (e.g. "Da gibt es eine Haltestelle." - "There is a tram stop here.").

Condition	Location	Motion
3D	36.81%	63.19%
2D	14.31%	85.69%

Table 2. Location and motion descriptions by instructor (means per trial).

In the 3d condition, the participants used temporal and spatial deictic terms more frequently than in the 2d condition (*jetzt* 'now' 3d: 7.3 occurrences per 100 utterances, 2d: 2.73. *hier* 'here' 3d: 2.21, 2d: 0.14).

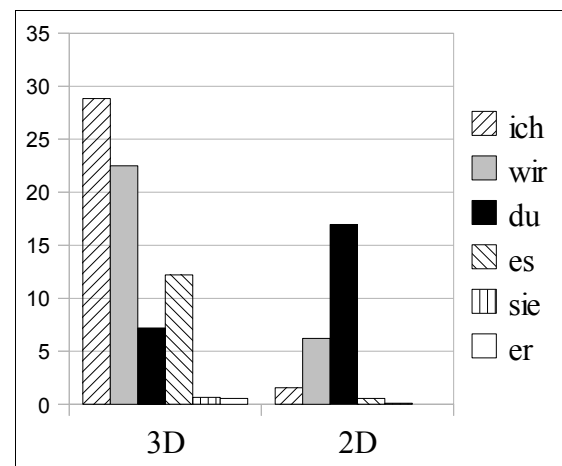


Figure 2. Relative frequency of personal pronouns in the two conditions.

5 Discussion

Our comparison of route directions given while perceiving an environment either as a 2d map or in a 3d view revealed that dimensionality has systematic consequences for discourse strategies on various levels. Location descriptions, route perspective expressions, 1st person singular personal pronouns, impersonal *es* 'it', as well as temporal and spatial deictic terms occurred more frequently in the instructors' discourse in the 3d condition than in the 2d condition. Also, in the 3d condition, instructors did not chunk route elements together. These findings reflect the fact that the instructors

consistently chose a different discourse strategy in this condition. Instead of producing procedural step-by-step instructions, they gave descriptions of the events happening to them and accounts of their surroundings, whereas instructors in the 2d condition gave typical route directions for their partner to follow.

There are three aspects that may be responsible for the different discourse strategies. First, it can be assumed that there is a habitual preference, due to the fact that people providing route directions usually have a 2d representation available to them, or prior knowledge of the relevant route, whereas someone navigating new surroundings would not normally be expected to provide efficient procedural instructions. Second, the lack of structural information in the 3d condition makes it difficult for instructors to describe the route from a survey perspective, or to deliver precise goal-oriented instructions. Third, in the 3d condition, progress for the instructor was slow - comparable to riding a bicycle along the route at moderate speed - due to the technical properties of Google Maps Street View. This severed the effect of the inherent lack of structural information, and most probably led the participants to verbalize their progress more frequently than necessary, in order to keep the conversation flowing, instead of waiting until they reached a point where more efficient instructions would be possible. This factor is also reflected in the number of utterances per trial: The higher number of utterances per trial in the 3d condition (see section 4) is at least partly a result of the technical setup.

In the case of chunking, time does not seem to be the only relevant factor. Klippel et al. (2003) showed that in a 2d scenario in which the route was only gradually revealed in the form of a moving dot on a map, participants still made use of chunking. It remains to be investigated whether the lack of chunking in the present scenario occurred due to the differing dimensionality, or resulted from the unconstrained real dialogue situation, in contrast to the pretend-dialogue used in Klippel et al. (2003).

Further research should differentiate the role of time in the choice of strategy from the impact of perspective. This requires experimental setups that allow for the systematic variation of the speed of the navigation, as well as for better control of such factors as previous knowledge and information density on the route. It

would also be necessary to examine two further conditions (instructor: 3d, follower: 2d and instructor: 2d, follower: 3d).

Acknowledgments

Funding by the DFG for the SFB/TR 8, project I5-[DiaSpace], is gratefully acknowledged. We thank the students who participated in our study, as well as Robert Porzel and Elena Andonova for their helpful advice.

References

- Juliana Goschler, Elena Andonova and Robert J. Ross. 2008. Perspective Use and Perspective Shift in Spatial Dialogue. In: Christian Freksa et al. (Eds.): *Spatial Cognition VI*, 250-265. Berlin: Springer.
- Alexander Klippel, Heike Tappe and Christopher Habel. 2003. Pictorial Representations of Routes: Chunking Route Segments during Comprehension. In: Christian Freksa et al. (Eds.): *Spatial Cognition III*, 11-33. Berlin: Springer.
- Robert E. Longacre. 1983. *The Grammar of Discourse*. Plenum Press, New York / London.
- Pierre-Emmanuel Michon and Michel Denis. 2001. When and Why Are Visual Landmarks Used in Giving Directions? In: Daniel R. Montello (Ed.): *COSIT 2001*, 292-305.
- Martin J. Pickering and Simon Garrod. 2004. Toward a Mechanistic Psychology of Dialogue. In *Behavioral and Brain Sciences*, 27:169-225.
- Karl Rehrl, Sven Leitinger, Georg Gartner and Felix Ortig. 2009. An Analysis of Direction and Motion Concepts in Verbal Descriptions of Route Choices. In: Kathleen Stewart Hornsby, Christophe Claramunt, Michel Denis and Gérard Ligozat (Eds.): *COSIT 2009*, 471-488. Springer-Verlag, Berlin / Heidelberg.
- Holly A. Taylor and Barbara Tversky. 1996. Perspective in Spatial Descriptions. In *Journal of Memory and Language*, 35:371-391.
- Matthew E. Watson, Martin J. Pickering and Holly P. Branigan. 2006. Why Dialogue Methods are Important for Investigating Spatial Language. In: Kenny R. Coventry, John Bateman and Thora Tenbrink (Eds.): *Spatial Language in Dialogue*, 8-22. Oxford University Press, Oxford.