Route Communication in Dialogue: a Matter of Principles

Theodora Koulouri
Department of Information Systems and Computing
Brunel University
Middlesex UB8 3PH
theodora.koulouri@brunel.ac.uk

Stanislao Lauria
Department of Information Systems and Computing
Brunel University
Middlesex UB8 3PH
stasha.lauria@brunel.ac.uk

Abstract

The present study uses the dialogue paradigm to explore route communication. It revolves around the analysis of a corpus of route instructions produced in real-time interaction with the follower. It explores the variation in forming route instructions and the factors that contribute in it. The results show that visual co-presence influences the performance, conversation patterns and configuration of instructions. Most importantly, the results suggest an analogy between the choices of instruction-givers and the communicative actions of their partners.

1.1 Spatial language in dialogue

The main question this paper attempts to address is how people produce route instructions in dialogue. The current zeitgeist in language research and dialogue system development seems to be the unified investigation of spatial language and dialogue (Coventry et al., 2009). Indicative of the growing prioritisation of dialogue in the study of spatial language are the on-going research efforts within the MapTask[1] project and the GIVE challenge[2] .

1.2 A framework for the analysis of route instructions

The study uses CORK (Communication of Route Knowledge, (Allen 2000)), a framework which provides a component-based analysis of route instructions. The CORK taxonomy differentiates between instructions that are directives (action statements with verbs of movement) and descriptive statements (with state-of-being verbs, like “be” and “see”). Descriptives present a static picture of spatial relations and provide the follower the opportunity to verify his position or reorient himself. The taxonomy also considers elements that provide specificity and distinguishing information about environmental features, called delimiters. Within this framework, Allen (2000) describes a set of principles pertaining to the configuration of route descriptions. Namely, people concentrate descriptives and delimiters on points along the route that offer for uncertainty (like crossroads). Moreover, the selection and placement of these components depends on the characteristics of the environment and the perceived needs of the follower. Evidence from empirical work supports the framework, reporting that errors in navigation increased when the route directions violated these principles. Nevertheless, the applicability of the suggested principles has only been tested in scenarios in which the directions were produced beforehand by either the experimenters or a separate group of subjects.

1.3 The effect of visual information

Studies exploring the effect of visual information on task-oriented interaction converge on that visual feedback leads to more efficient interactions and influences the conversational patterns between participants (Clark and Krych, 2004; Gergle et al., 2004; Koulouri and Lauria, 2009). These phenomena are generally attributed to the ease of establishing “common ground” when visual feedback is available. However, to the authors’ knowledge, most related studies have focused on high-level analysis of dialogue acts and many aspects of how interlocutors adapt their linguistic choices remain undefined.

1.4 Aims and hypotheses of study

The present study provides an empirical account of route instructions, as they emerge in real-time interaction with the follower. We offer the fol-

[1] http://www.hcrc.ed.ac.uk/maptask/
[2] http://www.give-challenge.org/research/

Proceedings of SIGDIAL 2010: the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 95–98,
The University of Tokyo, September 24-25, 2010.

lowing tentative hypotheses. Since visual co-presence facilitates grounding of information, it is expected to have a major effect on how route instructions are configured. Next, putting additional emphasis on the inter-individual processes involved in language use, this study aims to test whether the linguistic options mobilised by the instructor ultimately depend upon the contributions of the follower.

2 Methods

A study was designed to elicit natural route instructions in a restricted context. Pairs of participants collaborated in a navigation task, in a "Wizard-of-Oz" set-up. The instructors provided instructions to navigate their partners to designated locations in a simulated town, being under the impression that they were interacting with a software agent (robot). The study manipulated two factors; i) availability of visual information on follower's actions and ii), follower's interactive capacity. With regard to the first factor, there were two conditions in which the ability to monitor the actions of the "robot" was either removed or provided. The second factor also involved two conditions. In the first condition, the followers could interact using unconstrained language (henceforth, "Free" condition). In the second condition (henceforth, "Constrained" condition), a set of predetermined responses available to the followers aimed to coerce them towards more "automated" contributions; for instance, "open-ended" repairs such as "What?", which provide no specific information on the source of the problem. However, the followers were still able to be interactive if they wished so, by clicking the relevant buttons to request clarification or provide location information.

The study followed a between-subjects factorial design. A total of 56 students were allocated in the four conditions: Monitor-Free, Monitor-Constrained, No Monitor-Free, No Monitor-Constrained. The experimental procedure is described in detail in (Koulouri and Lauria, 2009).

2.1 Set-up

The experiment relied on a custom-built system which supported the interactive simulation and enabled real-time direct text communication between the pairs. The interfaces consisted of a graphical display and a dialogue box.

The interface of the instructor displayed the full map of the simulated town (Figure 1). On the upper right corner of the interface, there could be a small "monitor", in which the robot's immediate locality was displayed, but not the robot itself. The presence of the monitor feature depended on the experimental condition.

The followers' interface displayed a fraction of the map, the surroundings of the robot's current position. The robot was operated by the follower using the arrow keys on the keyboard. In the "Free" conditions followers could freely type messages. In the "Constrained" conditions, the followers needed to use the buttons on the interface (Figure 2).

Figure 1. The instructor's interface in the Monitor conditions. The monitor window on the upper right corner was removed in No Monitor conditions.

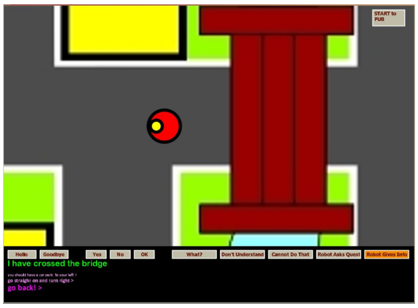

Figure 2. The follower's interface in the Constrained conditions. In the Free conditions, there were no buttons and followers could freely type any message.

2.2 Data analysis

The analysis of the corpus of route instructions followed the CORK framework (Allen, 2000). Communicative statements were classified as **Directives** or **Descriptives**. These communicative statements could contain references to environmental features. The types of environmental features considered were: **Locations** (e.g., buildings or bridges), **Pathways** (e.g., streets), **Choice Points** (e.g., junctions) and **Destination**. Last, instructions can be composed of delimiters, which fall into four categories:

1. **Distance designations:** e.g., "...*until* you see a car park".
2. **Direction designations:** e.g., "go *left"*.
3. **Relational terms:** e.g., *"on your left",*
4. **Modifiers:** e.g., "*big red* bridge", "take the *first/second/last* road".

3 Results

The experiment yielded a large corpus of 160 dialogues, composed of 3,386 turns. 1,485 instructions were collected. First, the analysis considers some common measures of efficiency. Next, the results of the component analysis of instructions are presented.

3.1 Efficiency of interaction

The number and length of turns and instructions and time needed to complete each task are typically used as measures of the efficiency of interaction. Additionally, fewer execution and understanding failures are taken as indicators of superior performance.

Time, number of turns, words and instructions: The ANOVA performed on time per task showed no reliable significant differences among groups. On the other hand, significant effects were observed with regard to all other dependent variables. An interaction effect was revealed after analysis on numbers of turns (F(1, 24) = 3.993, p = .05). Pairs in the Monitor-Free condition required less turns to complete the task compared to the other groups (column 1 of Table 1). It seems however that instructors in both Monitor conditions were dominating the conversational floor, having produced about 58% of the turns, compared to instructors in the No Monitor conditions (F(1, 24) = 5.303, p = .03). Nevertheless, it was not the case that instructors in Monitor conditions were "wordier". The number of words was similar among all instructor groups. The results indicated that the total number of words required to complete a task was much lower in Monitor conditions (F(1, 24) = = 5.215, p = 0.03) (see column 3 in Table 1). Next, instructors in Monitor conditions gave more instructions to guide the followers to the destination (F(1, 24) = 3.494, p = .07). However, these instructions were considerably shorter compared to the instructions provided by No Monitor instructors (F(1, 24) = 4.268, p = .05). All differences are amplified in the Monitor-Constrained group, in which more turns and instructions were needed but with fewer words and the "turn possession" of the instructor was the highest among the groups.

Condition	#Turns per task	%Instructor Turns	#Words per task	#Words per Instruction	#Instructions per task	Miscommunication per task
M-F	16.74	57.12%	87.33	4.70	9.08	1.14
M-C	23.95	58.86%	65.02	3.01	11.73	2.05
NM-F	23.63	52.28%	105.38	5.29	8.58	1.20
NM-C	20.15	50.62%	100.35	5.07	7.68	0.69

Table 1. Summary of Results (mean values).

Frequency of miscommunication: Miscommunication was calculated by considering two measures: the number of execution errors and of follower turns that were tagged as expressing non-understanding. The ANOVA revealed an interaction effect (F (1, 24) = 4.012, p = .05). Striking differences were observed between the Monitor-Constrained group and the rest; in particular, followers in this condition were twice or three times more likely to fail to understand and execute instructions (see last column in Table 1).

3.2 Component analysis of instructions

This section presents the results of the analysis on inclusion of landmark references, types of delimiters and communicative statements.

Landmark references: Instructors in both No Monitor conditions preferred to produce instructions that were anchored on landmarks, especially on 3D locations such as buildings (28% of instructions contained locations vs 14% in the Monitor conditions, (F (1, 24) = 12.034, p = .002)). On the other hand, Monitor instructors opted for simple action prescriptions. Particularly, 75% of the instructions in the Monitor-Constrained condition omitted any kind of reference (compared to an average of 42% in the other conditions).

Delimiters: Category 2 delimiters that provided simple direction information were prevalent in Monitor conditions (F (1, 24) = 11.407, p = .002). Further, an interaction effect was found (F (1, 24) = 3.802, p = .01); the number of category 2 delimiters almost doubles in the Monitor-Constrained condition. On the contrary, the use of category 1 delimiters, which provide information on the boundary of the route is very limited in the Monitor-Constrained condition (F (1, 24) = 5.350, p = .03). The third category of delimiters includes terms that specify the relation between traveller and an environmental feature ("on your left") or between environmental features. Again, the difference arises in the Monitor-Constrained condition, which included the lowest number of category 3 delimiters (marginal effect, F (1, 24) = 3.392, p = .07). Finally, the analysis performed on the frequency of category 4 delimiters did not yield any significant effect.

Directive and descriptive communicative statements: An interaction effect was revealed with regard to the proportion of directives and descriptives in the corpus (F (1, 24) = 3.830, p = .06). The instructors in the Monitor-Constrained condition produced less descriptives, which give information about relations among features in the environment and tap perceptual experience ("you will see a bridge"). In particular, in the Monitor-Constrained condition, 4.7% of instructions were descriptives, whereas the proportion of descriptives averaged 10% in all other conditions.

4 Discussion

The results resonate with previous research. The actions of the followers served as an immediate, accurate and effortless indicator of their current state of understanding, making verbal feedback redundant. Monitor instructors could readily confirm their assumptions about the information requirements of followers and used linguistics shortcuts and simpler instructions exactly at the moment needed. On the other hand, in the No Monitor condition, uncertainty about the position and movement of the robot created the need for elaborate and explicit instructions. The contribution of the present study lies on that it grounds these observations on quantitative analysis, using measures like words, turns and the relative frequencies of certain types of instruction components that vary in information value. Most importantly, it describes the specific ways in which instructors configure their directions in the presence/absence of visual information.

The CORK framework predicts that route protocols which are rich in descriptives and relational terms are associated with more successful navigation, compared to simple directional ones. Our results partially meet this expectation, since large numbers of execution errors and non-understandings were only observed in the Monitor-Constrained condition, whereas miscommunication rates were similar across the other groups. Indeed, this condition was found to generally differ from the rest. In particular, In the Monitor-Constrained condition, the dialogues were the shortest in terms of words. Instructors produced many but short instructions. The component-based analysis revealed that they employed overwhelmingly more action-based instructions without landmark references and descriptives. Boundary information on the route, frame of reference and spatial relations between environmental features were typically omitted. In both Constrained conditions, followers were expected to resort to a "mechanical" interaction, as coerced by the presence of the predefined responses. Inspection of the dialogues revealed that followers in the Monitor-Constrained condition did so, given the precedence of visual feedback. This was not the case with No Monitor-Constrained followers who needed to verbally ground information. Dialogue examples are provided in Table 2 below.

I: turn around I: go straight ahead I: stop I: turn left here I: go ahead F: *What?* I: Go straight ahead	I: Now keep going down the road until you see a car park F: *I am in front of the car park* I: turn right and walk till the end, along the road you will see a gym on your right F: *yes gym to my right side* I: good, keep going straight and you will see a factory on your left

Table 2. Dialogue excerpts from the Monitor-Constrained (column 1) and No Monitor-Constrained (column 2) conditions.

Thus, we propose that the linguistic choices of the followers "prime" the instructor's own strategies. In the Monitor-Constrained Condition, followers were less interactive, and gave fewer responses with lower information value. In harmony, their partners provided less elaborate instructions, which also lacked important information and specificity.

In conclusion, the findings confirm our initial hypotheses. Instructions are sensitive to conditions of (visual) co-presence. Moreover, a direct link was identified between the way in which instructions and follower's contributions are formulated. Following this lead, we are now focusing on a fine-grained analysis of the utterances of the follower.

References

Darren Gergle, Robert E. Kraut and Susan E. Fussell. 2004. Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. *Journal of Language and Social Psychology,* 23(4):491-517. Sage Publications, CA.

Gary L. Allen. 2000. Principles and Practices for Communicating Route Knowledge. *Applied Cog. Psychology.*14(4):333–359.

Herbert H. Clark and Meredyth A. Krych. 2004. Speaking While Monitoring Addressees for Understanding. *J. of Memory and Language*, 50:62-81.

Kenny Coventry, Thora Tenbrink and John Bateman, 2009. Spatial Language and Dialogue: Navigating the Domain. In K. Coventry, T. Tenbrink, and J. Bateman (Eds.) Spatial Language and Dialogue. 1-8. Oxford University Press. Oxford, UK.

Theodora Koulouri and Stanislao Lauria. 2009. Exploring Miscommunication and Collaborative Behaviour in Human-Robot Interaction, *SIGdial09*.